University of Alberta

Development of Mass Spectrometry for Bacteria Identification

by

Kevin Yoshio Dunlop

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

Department of Chemistry

Edmonton, Alberta

Spring 2001

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60423-3

Canada

University of Alberta

Library Release Form

Name of Author: Kevin Yoshio Dunlop

Title of Thesis: Development of Mass Spectrometry for Bacteria Identification

Degree: Master of Science

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Kevin Dunlop

148 Bestwick Dr.
Kamloops, BC
V2C 1M8

January 25, 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Development of Mass Spectrometry for Bacteria Identification submitted by Kevin Yoshio Dunlop in partial fulfillment of the requirements for the degree of Master of Science.

Dr. L. Li, Professor of Chemistry

Dr. C. Lucy, Professor of Chemistry

Dr. K. Froese, Assistant Professor of Environmental Health Sciences

Date: 1/29/01

to my parents

Herb and Shirley

ABSTRACT

Mass spectral analysis of large biological molecules such as proteins has only recently been employed for characterization of microorganisms. Suppression effects during ionization severely limited the number of signals observed through direct matrix-assisted laser desorption/ionization (MALDI) analysis of bacterial cell lysates. Fractionation of the lysate followed by MALDI, reveals the presence of an order of magnitude more components than direct MALDI. A reduction in suppression was observed through analysis with liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS). The online separation and mass spectral analysis produced approximately twice as many masses as direct MALDI. MALDI mass spectra of bacterial cell lysates at different stages of growth, exhibited substantial differences in the number of masses detected. However, LC/ESI-MS revealed much less dependence on culture time. In another study, the complex chromatogram produced by LC/ESI-MS was simplified for automated peak identification and mass spectral deconvolution. Removal of the HPLC static mixer resulted in an oscillation of the total ion current. The artificial peaks were found to facilitate automated data analysis and provided an alternative to the lengthy and operator dependant task of manually interpreting the results.

Acknowledgements

I would like to thank my supervisor Dr. Liang Li for his advice, support and encouragement throughout my project. I am grateful for the opportunity to have been a part of his research group. I also thank all members of Dr. Li's group with whom I have worked.

I thank the members of my examining committee, Dr. Charles Lucy and Dr. Kenneth Froese for their comments and suggestions regarding this thesis.

Finally, I thank the University of Alberta and the Department of Chemistry for a graduate teaching assistantship.

TABLE OF CONTENTS

CHAPTER ..... **PAGE**

1 **Introduction**

1.1 Identification of Biomarkers in Microorganisms 1

1.1.1 *Bacterial structure/classes* 2

1.1.2 *Classical methods* 3

1.1.3 *MALDI & ESI-MS* 4

1.1.3.1 *The MALDI process* 5

1.1.3.2 *The ESI process* 8

1.2 Liquid Chromatography/Electrospray Ionization MS 10

1.2.1 *Quadrupole mass analyzers* 11

1.2.2 *Ion-trap mass analyzers* 12

1.3 Nanoelectrospray Ionization MS 15

1.4 Brief Summary of Thesis 16

1.5 Literature Cited 16

2. **Nanospray Techniques and Applications**

2.1 Introduction 21

2.2 Experimental 22

2.2.1 *Materials* 22

2.2.2 *Needle fabrication* 23

2.2.3 *Cytochrome c tryptic digest* 24

2.3 Results and Discussion 24

2.3.1 *Background species* 24

2.3.2 *Imaging nanoelectrospray emitter tips* 29

2.3.3 *Cytochrome c tryptic digest* 30

2.3.4 *Fragmentation of highly charged bacterial protein* 33

2.4 Conclusions 37
2.5 Literature Cited 38

3 **Detection of Low-Mass Bacterial Proteins by LC/ESI-MS and MALDI**

3.1 Introduction 39
3.2 Experimental 41
3.2.1 *Materials* 41
3.2.2 *Sample preparation* 41
3.2.3 *Instrumentation* 42
3.3 Results and Discussion 44
3.3.1 *LC/ESI-MS* 44
3.3.2 *MALDI* 49
3.3.2.1 *Direct MALDI* 49
3.3.2.2 *Offline MALDI* 53
3.3.3 *Method evaluation* 56
3.3.3.1 *Comparison of 3 data sets* 59
3.3.3.2 *Database creation* 60
3.3.3.3 *Comparison with proteome database* 62
3.4 Conclusions 63
3.5 Literature Cited 64

4 **MALDI-TOF MS and LC/ESI-MS Analysis of *E. coli* 9637 Harvested at Different Growth Times**

4.1 Introduction 66
4.2 Experimental 67
4.2.1 *Materials* 67
4.2.2 *Sample Preparation* 67
4.3.3 *Instrumentation* 68
4.3 Results and Discussion 69
4.3.1 *MALDI* 69
4.3.2 *LC/ESI-MS* 73

4.4 Conclusions 80
4.5 Literature Cited 81

5. **Automated Analysis of LC/ESI-MS Data from Bacterial Extracts**

5.1 Introduction 82
5.2 Experimental 83
5.2.1 *Materials* 83
5.2.2 *Sample Preparation* 84
5.2.3 *Instrumentation* 84
5.3 Results and Discussion 86
5.3.1 *Automated peak identification* 88
5.3.1.1 *Manual data interpretation* 88
5.3.1.2 *Automated data interpretation* 89
5.3.2 *Manual Peak Identification* 91
5.3.2.1 *Manual data interpretation* 91
5.3.2.2 *Automated data interpretation* 92
5.4 Conclusions 99
5.5 Literature Cited 100

6. **Conclusions and Future Work**

6.1 Conclusions & Future Work 101

LIST OF TABLES

Table 3-1 Optimized separation conditions (shown as % ACN) 43

Table 3-2 $(M+H)^+$ from *E. coli* extract by online LC/ESI-MS 48

Table 3-3 $(M+H)^+$ from *B. megaterium* extract by online LC/ESI-MS 48

Table 3-4 $(M+H)^+$ from *C. freundii* extract by online LC/ESI-MS 48

Table 3-5 $(M+H)^+$ from *E. coli* crude mixture by direct MALDI 51

Table 3-6 $(M+H)^+$ from *B. megaterium* crude extract by direct MALDI 51

Table 3-7 $(M+H)^+$ from *C. freundii* crude extract by direct MALDI 51

Table 3-8 $(M+H)^+$ from *E. coli* extract by LC/offline MALDI 54

Table 3-9 $(M+H)^+$ from *B. megaterium* extract by LC/offline MALDI 55

Table 3-10 $(M+H)^+$ from *C. freundii* extract by LC/off-line MALDI 56

Table 4-1 MALDI $(M+H)^+$ 6 h incubation 72

Table 4-2 MALDI $(M+H)^+$ 8 h incubation 72

Table 4-3 MALDI $(M+H)^+$ 10 h incubation 73

Table 4-4 MALDI $(M+H)^+$ 12 h incubation 73

Table 4-5 LC/ESI-MS $(M+H)^+$ 6 h incubation 76

Table 4-6 LC/ESI-MS $(M+H)^+$ 8 h incubation 77

Table 4-7 LC/ESI-MS $(M+H)^+$ 10 h incubation 77

Table 4-8 LC/ESI-MS $(M+H)^+$ 12 h incubation 77

Table 5-1 **Automated peak identification.** Masses found through both manual and automated interpretation of the deconvoluted mass spectra by ChemStation. The masses listed were found by deconvoluting mass spectra from peaks found automatically through removal of the static mixer. At the bottom of the columns are listed the number of counts in each mass spectrum below which potential peaks are rejected. 90

Table 5-2 **Manual peak identification.** Masses found through both manual and automated interpretation of the deconvoluted mass spectra by ChemStation. The masses listed were found by deconvoluting mass spectra from peaks found by averaging over the listed time frame. 93

LIST OF FIGURES

Figure 1-1 Diagrams representing the A) gram-positive and B) gram-negative cell wall. 3

Figure 1-2 Schematic of a MALDI mass spectrometer showing the ion source and drift region where ions are separated according to their flight time to the detector. 7

Figure 1-3 Schematic showing introduction of HPLC effluent into the HP1100 MSD quadrupole mass spectrometer 9

Figure 1-4 Configuration of a quadrupole mass analyzer showing ion introduction and the stable trajectory of an ion between the four electrodes. 11

Figure 1-5 Diagram of an ion-trap showing entrance of ions of various m/z ratios, their confinement within the trap, and subsequent ejection from the smallest to largest m/z ratio 13

Figure 2-1 Nanospray MS^2 spectrum of 1522.1 m/z component of the Hewlett Packard ES tuning mix for Esquire-LC standard mass range. 25

Figure 2-2 (A) Background species present when spraying acetonitrile in the nanoelectrospray emitters fabricated from aluminosilicate glass. (B) Background species present after baking the emitters at 218 °C for 20 hours 26

Figure 2-3 (A) MS^2 spectrum of the 571.6 ion from Figure 2-1
(B) MS^3 spectrum of the 300.4 ion from (A) 27

Figure 2-4 (A) MS^2 spectrum of the 543.5 ion from Figure 2-2
(B) MS^3 spectrum of the 272.3 ion from (A) 28

Figure 2-5 Nanoelectrospray emitter pulled from aluminosilicate glass from Sutter Instrument Co. (1.2 mm o.d., 0.85 mm i.d., catalog # AF120-85-10). Tip was pulled with a P-97 Flaming/Brown micropipette puller with program described in 2.2.2. 29

Figure 2-6 Nanoelectrospray spectrum of a 1 μM equine cytochrome c solution dissolved in water/methanol/acetic acid at 50/48/2 (v/v/v). 30

Figure 2-7 Tryptic peptide map of equine cytochrome c produced using the fabricated nanoelectrospray emitters. Labeled masses correspond to cytochrome c peptides. Final protein concentration was 19 μM dissolved in water/methanol/acetic acid to 50/49/1 (v/v/v). 31

Figure 2-8 MS^2 of 584.9 peak from Figure 2-7 The doubly charged $M+2H^+$ species produces many singly charged products at higher m/z ratios. The fragments correspond to an amino acid sequence of TGPNLHGLFGR. 32

Figure 2-9 MS^2 of 907.5 peak from Figure 2-7. The fragments correspond to an amino acid sequence of MIFAGIKK. 33

Figure 2-10 (A) Principal component in fraction 60 obtained from LC/ESI-MS $(M+H)^+ = 9226.0$ Da
(B) Fraction 60 analyzed by nanoelectrospray MS. 35

Figure 2-11 MS/MS spectrum of the +11 charge state (m/z = 839.9) Figure 2-10 (B)..... 36

Figure 3-1 Effect of increasing fragmentation voltage on charge state envelope. A higher voltage drop in the CID region (see Figure 1-2 for instrument schematic) strips protons from the protein and results in a shift to lower charge states. (D) shows the wider charge envelope and increase in signal observed when the fragmentation voltage is ramped. 46

Figure 3-2 Total ion chromatograms observed with static mixer removed..... 47
(A) *E. coli*
(B) *B. megaterium*
(C) *C. freundii*

Figure 3-3 Direct MALDI spectra of (A) *E. coli*, (B) *B. megaterium* and (C) *C. freundii* crude extract..... 50

Figure 3-4 Direct MALDI spectra of *E. coli* crude extract in different extraction volumes..... 52
(A) 1.5 mg *E. coli* in 1.5 mL 0.1% (v/v) TFA
(B) 1.5 mg *E. coli* in 200 µL 0.1% (v/v) TFA

Figure 3-5 Comparison of species detected in fraction 39. (A) shows peaks observed by MALDI. (B) shows the smaller number seen by LC/ESI-MS. A total of six species were found; however, only five are shown for clarity. 58

Figure 4-1 Extraction efficiency shown here with MALDI mass spectra of *E. coli* 9637 (10 hr. incubation) extracts. The first 1 mL extract is shown on the bottom, followed by each successive 1 mL extraction. 70

Figure 4-2 MALDI mass spectra of *E. coli* 9637 bacterial extracts at different incubation times. The concentrated raw extracts were mixed 1:2 with saturated HCCA in formic acid / methanol / water (1/2/3) (v/v/v)..... 71

Figure 4-3 Total ion chromatograms of *E. coli* 9637 bacterial extracts at different incubation times. Glacial acetic acid was added post-column at 100 µL/min. 74

Figure 4-4 UV chromatograms at 214 nm of *E. coli* 9637 extracts at different incubation times. Both the UV and MS detectors were connected in series with the effluent passing first through the UV and then the MS detector..... 75

Figure 4-5 Total ion chromatograms of *E. coli* 9637 (10 hr. incubation) at varying concentrations of post-column acetic acid..... 79

Figure 4-6 Mass spectral signal enhancement observed upon addition of post column acetic acid at the listed concentrations. Acid was added at 100 μL/min, or one half the mobile phase flow rate..... 80

Figure 5-1 TIC of *E. coli* 9637 cell extract with post-column glacial acetic acid. 87
(A) Static mixer removed
(B) Static mixer installed

Figure 5-2 TIC of *B. globigii* cell extract with post-column glacial acetic acid. .. 96
(A) Static mixer removed
(B) Static mixer installed

Figure 5-3 Example of mass spectra showing charge states observed from *E. coli* cell extracts. The majority of the peaks, both manually and automatically identified, contained more than one protein.
(A) Two components with overlapping charge envelopes
A. $(M+H)^+ = 11186$
B. $(M+H)^+ = 18162$
(B) One component with $(M+H)^+ = 7707.5$..... 98

LIST OF ABBREVIATIONS

ATCC	American Type Culture Collection
B. globigii	*Bacillus globigii*
B. subtilis	*Bacillus subtilis*
CID	collision induced dissociation
C. freundii	*Citrobacter freundii*
ESI	electrospray ionization
E. coli	*Escherichia coli*
HCCA	α-cyano-4-hydroxycinnamic acid
HPLC	high performance liquid chromatography
LC/ESI-MS	liquid chromatography/electrospray ionization mass spectrometry
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
MS^n	tandem mass spectrometry
PEEK	poly(ether ether ketone)
RF	radio frequency
TFA	trifluroacetic acid
TIC	total ion chromatogram
TOF	time-of-flight
UV	ultraviolet
v/v	volume-to-volume ratio

CHAPTER 1

Introduction

1.1 Identification of Biomarkers in Microorganisms

The characterization of microorganisms is important in matters of food production, public health and in recognizing potential biological and environmental hazards [1]. Developing methods that are fast, straightforward, efficient and offer a high degree of certainty, is crucial for complementing the current techniques for microbial classification. Identification of biomarkers, or unique chemical constituents, is one approach for classifying microorganisms. Several cellular components have been examined as possible markers, including lipids [2,3], carbohydrates [4], DNA [5] and proteins [6].

With the advent of mass spectrometric ionization techniques that enable ionization of intact biological samples, bacterial proteins are becoming a major focus of study for characterizing bacteria. An average bacterial cell contains approximately 10% lipid and more than 50% protein [7]. As a result, proteins exhibit great potential for use as markers for characterizing bacteria. The wealth of information obtained through analysis of eukaryotic and prokaryotic organisms, is being catalogued and is available for searching in databases containing both gene and protein amino acid sequences [8,9]. There are currently several databases available through the Internet, which contain the theoretical proteins that may be expressed by microorganisms. Searching the online SWISS-PROT and TrEMBL database with the Sequence Retrieval System, the microorganism with the largest completed database is *Escherichia coli*, with a total of

7478 proteins [10]. According to the National Centre for Biotechnology Information, complete genomes of 18 microorganisms are now known [11].

One approach to bacterial identification involves detecting and cataloguing the masses of proteins expressed by bacteria. These protein masses would then be searched against the public databases and potential matches would be retrieved [12]. Characteristic proteins could then be used to identify bacteria according to their genus, species or even strain [13]. An alternative approach relies on producing mass spectral fingerprints of protein mass spectra, and then comparing unknown spectra to the archived sets [14,15]. This approach however, suffers from a lack of mass spectral reproducibility even in cases where identical cultures are used in replicate experiments.

1.1.1 Bacterial structure / classes

Bacteria, or prokaryotes, are commonly divided into two groups on the basis of staining with crystal violet and iodine. Cells retaining the crystal violet/iodine complex in their walls after washing with ethanol are deemed gram-positive. In cases where the dye is removed, safranin dyes the remaining cells pink, which are then classified gram-negative.

Both gram-positive and gram-negative bacteria possess peptidoglycan in their cell walls; however, gram-positive cells are richer in peptidoglycan and so have much more rigid cell walls. Figure 1-1 shows the structure of the two classes of cell walls.

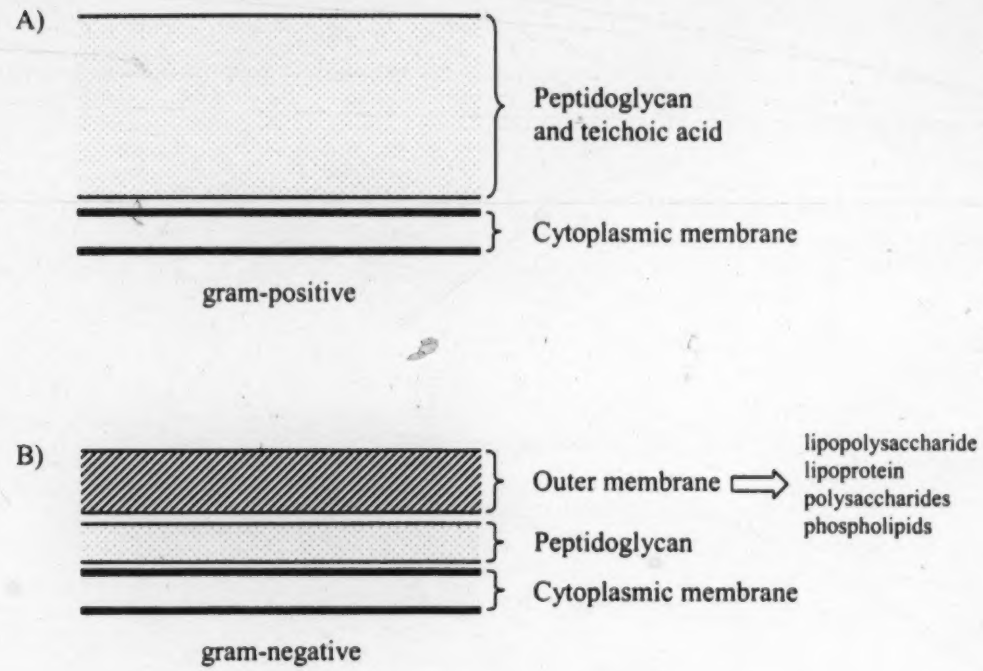

Figure 1-1 Diagrams representing the A) gram-positive and B) gram-negative cell wall

1.1.2 Classical Methods

Traditional methods for characterizing bacteria are based on phenotypic properties. Such methods group strains according to certain characteristics such as staining properties, morphology, motility, spore formation and pigmentation [16]. Examples of such methods include phage typing [17], bacteriocin typing [18], serotyping [19] and biotyping [20].

More recent developments in bacterial identification and classification are based on molecular techniques and are known as genotypic typing methods. These methods examine the genetic makeup of a microorganism [21]. These techniques focus on

separating bacterial DNA fragments or plasmids by electrophoresis [22] or discriminating cells by flow cytometry [23].

Several detection methods other than mass spectrometry also exist for characterizing bacteria. Some of these methods include bioluminescence detection of adenosine triphosphate content in bio-aerosols [24], sodium dodecyl sulfate-polyacrylamide gel electrophoresis of bacterial proteins [25], infrared detection of bio-aerosols [26], and gas chromatography of cell wall hydroxy fatty acids [27].

1.1.3 MALDI & ESI-MS

A mass spectrometer measures the mass to charge ratio (m/z) of molecules, where m is the atomic mass in Daltons and z is the number of elementary charge units. The process is accomplished through three steps: ionization, mass separation and detection. Prior to the 1980s, mass spectrometry was a tool used primarily for analysis of volatile organic compounds. Sample ionization through bombardment with high-energy electrons was the most commonly used ionization method available at the time, and was not well suited for producing intact gas phase biomolecules. In the late 1980s, two major advances were made in generating intact gas phase ions of biological origin. The two techniques, matrix-assisted laser desorption/ionization (MALDI), and electrospray ionization (ESI), revolutionized the investigation of biological materials.

In 1917, Zeleny [28] first reported the process of producing fine aerosols of charged droplets using an electric field. Many years later, Dole [29] used the electrospray process to generate gas phase polymeric ions; however, the technique focussed on a very narrow set of chemical species. It was Yamashita and Fenn [30]

who first coupled the electrospray ionization process to a quadrupole mass spectrometer. A group of Russian researchers published an independent report at the same time; however, their approach used a magnetic sector analyzer [31].

Hillenkamp and co-workers first described MALDI as a technique for the analysis of biological material in 1988 [32]. MALDI and ESI hold great potential for characterization of bacteria since genetic differences lead to variation in protein expression, which can then be studied through molecular weight measurements and peptide sequence analysis.

1.1.3.1 The MALDI process

The MALDI process [33,34] involves a transfer of energy from photons to the sample, which has been mixed with an organic matrix. The analyte of interest is mixed with an excess of suitable matrix compound and approximately 1 μL is deposited on a stainless steel surface. The most common matrices for analysis of peptides and proteins are: 2,5-dihydroxybenzoic acid, trans-3,5-dimethyloxy-4-hydroxycinnamic acid, 4-hydroxy-α-cyano cinnamic acid and 2-(4-hydroxyphenylazo)-benzoic acid. The surface is then irradiated with a laser, most frequently a 337 nm nitrogen laser, and the process of ionization is started. Once the matrix absorbs the laser energy, the analyte is lifted into the gas phase by the expanding plume of matrix molecules. Many models have been proposed to explain the mechanism of ion formation in MALDI [35-37]; however, there is no consensus on the exact sequence of events leading to the production of protonated/deprotonated gas phase ions. One study proposes extraction of protons from the excited matrix and their subsequent attachment to the analyte [38].

The MALDI process produces gas phase ions in a very gentle manner, and as a result, is perfectly suited for ionizing large, nonvolatile biomolecules such as peptides, proteins, oligonucleotides and oligosaccharides. Spectra are characterized by low charge states, contrasting greatly with that produced by ESI. In addition, MALDI is known for its ability to tolerate samples contaminated with salts, buffers and detergents; however, alkali and matrix adducts are still frequently observed. The presence of such adducts and matrix clusters can be greatly diminished through the use of on-probe washing [39].

The discontinuous nature of the MALDI process favours a pulsed analyzer and the most common type of detector is a time-of-flight (TOF). Figure 1-1 shows an example of a typical MALDI source and TOF analyzer. After the laser pulse strikes the sample plate, a packet of analyte molecules is desorbed into the gas phase and accelerated to a fixed kinetic energy by an electric field. During their flight through the field free region of the flight tube, ions are separated into packets traveling with a velocity characteristic of their mass. Ions are separated according to their mass-to-charge ratio during their flight and a mass spectrum is produced based on their flight time. The TOF analyzer produces a complete mass spectrum for each measurement cycle in addition to having a simple design and a theoretically unlimited mass range. Increased resolution has been accomplished through development of delayed extraction and reflectron designs [40].

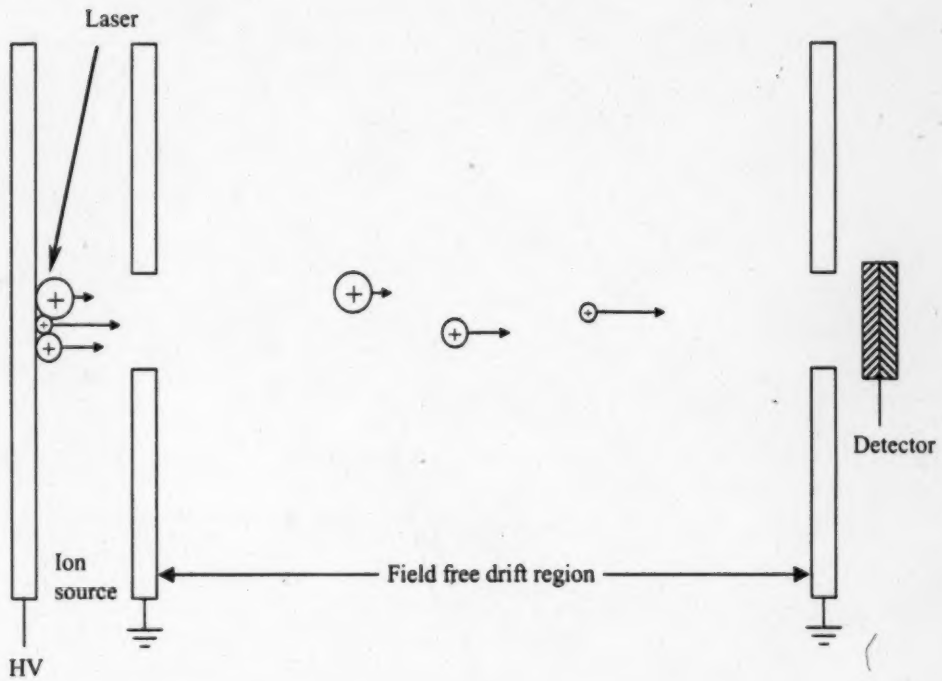

Figure 1-2 Schematic of a MALDI mass spectrometer showing the ion source and drift region where ions are separated according to their flight time to the detector.

1.1.3.2 The ESI process

The electrospray ionization process [41] involves flowing a solution of dissolved ions through a needle held at high potential (4-5 kV) at atmospheric pressure. The electric field produced at the needle tip (inner diameter of approximately 250 μm [42]) produces a fine aerosol of highly charged droplets. As the droplets shrink through solvent evaporation and disintegrations, highly charged gas phase ions are produced which are subsequently sampled by the mass spectrometer. The transfer of ions from solution to the gas phase imparts negligible internal energy to the analytes, and as a result, is perfectly suited for ionizing thermally labile molecules such as DNA and proteins. Figure 1-3 shows the setup of the electrospray needle in relation to the quadrupole analyzer that was used for the liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS) experiments described in this thesis.

The presence of multiply charged peaks is a striking feature of spectra produced via the ESI process. The extremely high charge density in the droplets produces gas phase ions with numerous charges associated with basic or acidic sites. This multiple charging phenomenon is beneficial since it effectively extends the mass range of any analyzer by a factor equal to the number of charges associated with the analyte. The multiple charging phenomenon results in an improvement in mass accuracy because each charge state yields an independent mass measurement, which can then be averaged over the entire charge envelope. However, the large number of peaks in the mass spectrum can quickly become difficult to interpret, in addition to placing a large burden on the resolution of the mass analyzer.

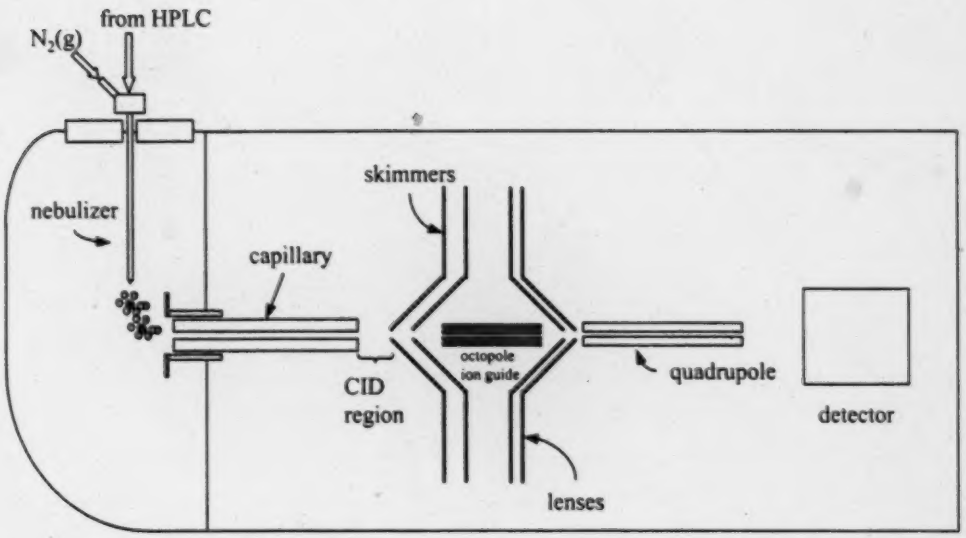

Figure 1-3 Schematic showing introduction of HPLC effluent into the HP1100 MSD quadrupole mass spectrometer.

1.2 Liquid Chromatography / Electrospray Ionization MS

The continuous ionization process that ESI provides leads naturally to its coupling with liquid separation methods such as liquid chromatography and capillary electrophoresis. A wide variety of ESI source designs are available which can manage flow rates ranging from hundreds of nL/min to 1-2 mL/min [43,44]. Flows in the hundreds of nL/min flow regime function as regular ESI systems; however, greater flow rates require assistance through nebulization and thermal techniques. These so called pneumatically assisted sources use a coaxial gas (usually nitrogen) to assist aerosol production, as well as heated nitrogen drying gas to improve analyte desolvation and declustering.

The analysis of complex mixtures is aided considerably by the chromatography employed before the actual ionization process. For extremely complex mixtures, such as biological protein extracts, the chromatography may not be sufficient to completely separate the components in the mixture. Complete separation however, is not always required since the ESI detection provides molecular weight and structural information. The chromatography also acts as an on-line decontamination step by removing low molecular weight species from the mixture in addition to salts, which may form adducts with the analytes of interest.

In order to improve the separation efficiency and resolution of complex mixtures, mobile phase modifiers are typically added to reverse phase HPLC separations. The use ion pairing agents like trifluoroacetic acid (TFA) however, strongly suppresses the formation of detectable gas phase peptides and proteins in ESI [45]. One way of dealing with such signal suppression is to infuse a weak acid after

separation and prior to ionization. Post-column addition of acetic, propionic, butyric, formic and valeric acids have been employed [46]. The introduction of these acids breaks up the ion pairs formed between the TFA anion and basic residues on the analyte.

Two common mass analyzers, which are coupled to the ESI source, are quadrupoles and ion traps.

1.2.1 Quadrupole mass analyzers

A quadrupole mass analyzer, or more appropriately mass filter, consists of a set of four cylindrical rods in which opposing rods are electrically connected [47]. Figure 1-4 shows a schematic representation of the quadrupole mass filter. One set of rods is held at a positive DC potential, and the opposing set is held at negative. Superimposed on the DC voltages is a set of RF voltages applied to each set of rods, with each pair being 180° out of phase.

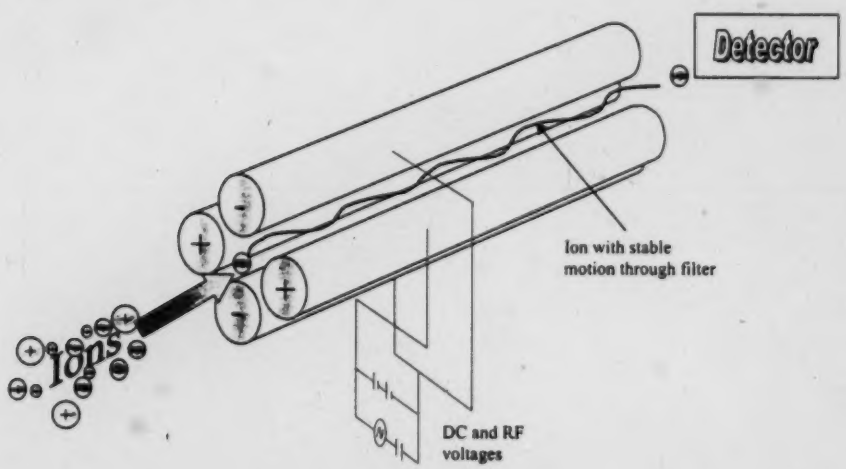

Figure 1-4 Configuration of a quadrupole mass analyzer showing ion introduction and the stable trajectory of an ion between the four electrodes.

This particular mass analyzer is best described as a filter, since it functions by selectively allowing ions within a mass window to pass through to the detector. Only those ions that are stable in both the XZ and YZ planes possess the motion required to pass between the rods and strike the detector. Ions that are unstable in the electric field are removed through collisions with the metal electrodes.

Successful passage of ions through the quadrupole mass filter is described by the solutions to the second order differential equation described by Mathieu [48]. Sweeping the RF and DC voltages while keeping their ratios constant results in filtering ions according to increasing m/z ratio.

1.2.2 Ion-trap mass analyzers

Instruments incorporating the quadrupole ion trap are based on the Paul trap design [49]. The quadrupole ion trap [50] is a versatile device performing two functions, namely storage of gaseous ions and selectively ejecting these ions thereby functioning as a mass spectrometer.

Early designs produced ions inside the trap by electron impact or chemical ionization [51]. This approach however, is unsuitable for studying large, thermally labile species of biological origin. More recent designs create ions externally and subsequently trap and scan according to their m/z ratio. The addition of helium into the trap (1-2 x 10^{-5} mbar) helps remove energy from the ions entering the trap, and helps confine them to the centre.

A diagram of an ion trap mass spectrometer is shown in Fig 1-5. The design consists of three electrodes with hyperbolic geometry, namely a ring electrode and two

end cap electrodes. Instead of filtering ions through, as the quadrupole mass filter in section 1.2.1, the quadrupole ion trap confines charged particles within an electric field. In one method of operation, a quadrupolar electric field is produced inside the trap through application of a radio frequency (RF) potential to the ring electrode, while the two end cap electrodes are held at ground. The 3D quadrupolar potential field produced inside the device traps the ions in a stable oscillating trajectory.

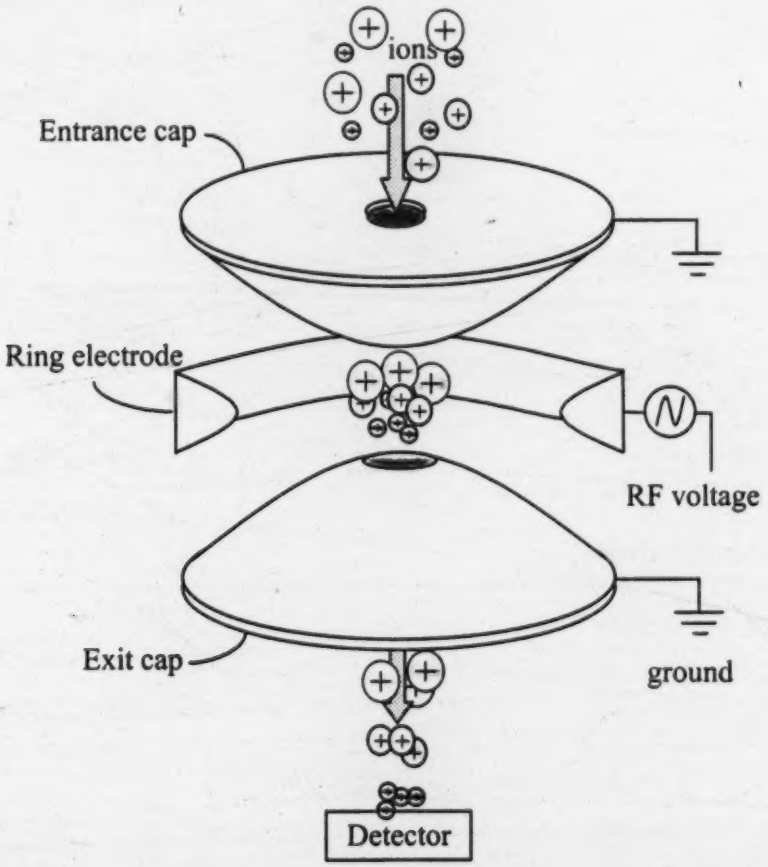

Figure 1-5 Diagram of an ion-trap showing entrance of ions of various m/z ratios, their confinement within the trap, and subsequent ejection from the smallest to largest m/z ratio. Adapted from [52].

A recent theoretical model has been proposed describing the trapping efficiency as a function of several instrument parameters such as phase of RF voltage and the energy of ions injected into the trap [52]. Trapped ions have a stable trajectory in both radial and axial directions; however, by ramping the RF voltage on the ring electrode, ions are made unstable in the axial direction and are ejected from the trap in order of increasing m/z. The stability and instability of ions within the confines of the applied electric field, is described by the solutions to the second-order differential equation described by Mathieu [47]. A detailed description of the Mathieu diagram and how it relates to ion motion within the trap is given in [53].

An alternative approach to scanning ions from the trap involves applying a supplementary AC voltage across the end-cap electrodes in a process called resonant ejection [54]. In this mode of operation, ions are ejected from the trap even though they are in the stable region of the Mathieu stability diagram. The ions are ejected once their secular frequencies match the frequency of the supplementary voltage applied to the end-caps.

The ion trap provides the user with the option of performing tandem MS analysis of trapped ions (MS^n). Under such operation, a wideband composite of frequencies is applied to the ring electrode, resulting in ejection of all ions except for the one of interest. The isolated ion can then be excited away from the centre of the trap. Collisions with helium buffer gas increase the internal energy of the ions, whereby they may undergo collision-induced dissociation (CID). Significant structural information can be obtained when operating the trap in such a manner.

1.3 Nanoelectrospray Ionization MS

A nanoelectrospray source is essentially a miniaturized version of the standard electrospray source. Since a report in 1996 [55], nanoelectrospray has become a powerful weapon in the mass spectrometric arsenal for peptide and protein identification. When confronted with limited analyte, the nanoelectrospray source allows one the luxury of extremely low flow rates (20 – 40 nL/min), which result in smaller droplet diameters when compared to the standard electrospray set-up (< 200 nm versus 1-2 μm) [17]. In addition to a dramatic increase in analysis time, nanoelectrospray offers more efficient sample consumption and increased sensitivity. Picolitre flow rates with fused-silica capillary tubing with an internal tip diameter of 2 μm have also been reported [56]. Theoretically only a few hundred nL of sample is required; however, handling such small volumes with regular laboratory pipettes is difficult. Moreover, air trapped in the tip can cause the flow to stop. As a result, volumes of at least 2 μL are easier to work with.

Conventional nanoelectrospray emitters are fabricated from borosilicate or aluminosilicate glass. The glass is heated and pulled to an extremely fine taper with an inner diameter from 1 to 10 μm. In order to provide electrical contact, the glass is either sputtered with gold [54], or a platinum electrode is inserted into analyte solution [57]. The electric field created between the emitter and the counter electrode is sufficient to disperse the analyte liquid thereby eliminating the need for pumps.

1.4 Brief Summary of the Thesis

MALDI and LC/ESI-MS detection of bacterial proteins is demonstrated in this thesis. Chapter 2 examines some technical aspects of fabricating nanoelectrospray emitters, as well as providing some examples of using the nanoelectrospray technique. Chapter 3 deals with some of the fundamental issues involved in using LC/ESI-MS and MALDI as tools for measuring bacterial protein masses. In Chapter 4, the effect of bacterial growth time on the number of masses detected is studied. Chapter 5 evaluates the potential for automated analysis of data obtained from LC/ESI-MS analysis of complex mixtures. Finally, conclusions and future work are presented in Chapter 6.

1.5 Literature Cited

[1] Welham, K. J.; Domin, M. A.; Scannell, D. E.; Cohen, E.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **1998**, 12, 176-180.

[2] Wiese, A.; Seydel, U. *J. Industrl. Microbiol. Biotech.* **1999**, 23, 414-424.

[3] Snyder, A. P.; Smith, B. W. P.; Dworzanski, J. P.; Meuzelaar, H. L. C., In *Mass Spectrometry for Characterization of Mircoorganisms*, ACS Symposium Series, vol. 541, C. Fenselau (Ed.), American Chemical Society, Washington, DC (**1994**) 62-84.

[4] Welham, K. J.; Domin, M. A.; Johnson, K.; Jones, L.; Ashton, D. S. *Rapid Commun. Mass Spectrom.* **2000**, 14, 307-310.

[5] Cox, N.; Johnston, J.; Szarka, Z.; Wright, D.; Archard, L. *J. Gen. Microbiol.* **1990**, 136, 1639-1643.

[6] Cash, P. *Anal. Chim. Acta*, **1998**, 372, 121-146.

[7] Cain, T. C.; Lubman, D. M.; Weber, W. J. *Rapid Commun. Mass Spectrom.* **1994**, 8, 1026-1030.

[8] Yates, J. R. *Electrophoresis*, **1998**, 19, 893-900.

[9] Beranova-Giorgianni, S.; Desiderio, D. M. *Rapid Commun. Mass Spectrom.* **2000**, 14, 161-167.

[10] http://expasy.cbr.nrc.ca/srs5/

[11] http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html

[12] Demirev, P.; Ho, Y.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, 71, 2732-2738.

[13] Krishnamurthy, T.; Ross, P. L. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1992-1996.

[14] Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1227-1232.

[15] Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1998**, 12, 630-636.

[16] Towner, K. J.; Cockayne, A. *"Molecular Methods for Microbial Identification and Typing"*, **1993**, Chapman & Hall, London, UK, 1.

[17] Toth, I. K.; Bertheau, Y.; Hyman, L. J.; Laplaze, L.; Lopez, M. M.; McNicol, J.; Niepold, F.; Persson, P.; Salmond, G. P. C.; Sletten, A.; van der Wolf, J. M.; Perombelon, M. C. M. *J. Appl. Microbiol.* **1999**, 87, 770-781.

[18] Avelar, K. E. S.; Pinto, L. J. F.; Antunes, L. C. M.; Lobo, L. A.; Bastos, M. C. F.; Domingues, R. M. C. P.; de Souza Ferreira, M. C. *Lttrs. Appl. Microbiol.* **1999**, 29, 264-268.

[19] Baudart, J.; Lemarchand, K.; Brisabois, A.; Lebaron, P. *Appl. Environ. Microbiol.* **2000**, 66, 1544-1552.

[20] Connor, K. M.; Quirie, M. M.; Baird, G.; Donachie, W. *J. Clin. Microbiol.* **2000**, 38, 2633-2637.

[21] Maslow, J.; Mulligan, M. E. *Infect. Control. Hosp. Epiemiol.* **1996**, 17, 595-604.

[22] Pfaller, M. A. *Arch. Pathol. Lab. Med.* **1999**, 123, 1007-1010.

[23] Kim, Y.; Jett, J. H.; Larson, E. J.; Penttila, J. R.; Marrone, B. L.; Keller, R. A. *Cytometry* **1999**, 36, 324-332.

[24] Stopa, P. J.; Tieman, D.; Coon, P.; Milton, M. M.; Paterno, D. *Field Anal. Chem. Tech.* **1999**, 3, 282-290.

[25] Schmid, R.; Bernhardt, J.; Antelmann, H.; Volker, A.; Mach, H.; Volker, U.; Hecker, M. *Microbiol.* **1997**, 143, 991-998.

[26] Gittins, C.; Piper, L.; Rawlins, W.; Marinelli, W.; Jensen, J.; Akinyemi, A. *Field Anal. Chem. Tech.* **1999**, 3, 274-282.

[27] Stead, D. E.; Sellwood, J. E.; Wilson, J.; Viney, I. *J. Appl. Bacteriol.* **1992**, 72, 315-321.

[28] Zeleny, *J. Phys. Rev.* **1917**, 10, 1-6.

[29] Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, 49, 2240-2249.

[30] Yamashita, M.; Fenn, J. *J. Phys. Chem.* **1984**, 88, 4451-4459.

[31] Aleksandrov, M. L.; Gall, L. N.; Krasnov, N. V.; Nikolaev, V. I.; Shkurov, V. A. *Zhur. Anal. Khim.* **1985**, 40, 1570-1580.

[32] Karas, M.; Bachman, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Proc.* **1987**, 78, 53-68.

[33] Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998**, 17, 337-366.

[34] Karas, M.; Gluckmann, M.; Schafer, J. *J. Mass Spectrom.* **2000**, 35, 1-12.

[35] Ehring, H.; Karas, M.; Hillenkamp, F. *Org. Mass Spectrom.* **1992**, 27, 472-480.

[36] Lehmann, E.; Knochenmuss, R.; Zenobi, R. *Rapid Commun. Mass Spectrom.* **1997**, 11, 1483-1492.

[37] Burton, R.; Watson, C.; Eyler, J.; Lang, L.; Powell, D.; Avery, M. *Rapid Commun. Mass Spectrom.* **1997**, 11, 443-446.

[38] Calba, P.; Muller, J.; Inouye, M. *Rapid Commun. Mass Spectrom.* **1998**, 12, 1727-1731.

[39] Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskaya, E.; Kroll-Kristensen, K.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997**, 32, 593-601.

[40] Weickhardt, C.; Moritz, F.; Jurgen, G. *Mass Spectrom. Rev.* **1996**, 15, 139-162.

[41] Kebarle, P.; Ho, Y. *Electrospray Ionization Mass Spectrometry*, R. B. Cole ed. **1997**, 3-63.

[42] Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M. Udseth, H. R. *Mass Spectrom. Rev.* **1991**, 10, 359-451.

[43] Niessen, W. M. A.; Tinke, A. P. *J. Chromatogr. A* **1995**, 703, 37-57.

[44] Abian, J.; Oosterkamp, A. J.; Gelpi, E. *J. Mass Spectrom.* **1999**, 34, 244-254

[45] Kuhlmann, F. E.; Apffel, A.; Fischer, S.; Goldberg, G.; Goodley, P. C. *J. Am. Chem. Soc.* **1995**, 6, 1221-1225.

[46] Apffel, A.; Fischer, S.; Goldberg, G.; Goodley, P.; Kutilmann, F. E. *J. Chromatogr. A* **1995**, 712, 177-190.

[47] Dawson, P. H. "*Quadrupole Mass Spectrometry and its Applications*", Elsevier: New York, **1976**.

[48] Mathieu, E. *J. Math. Pure Appl.* **1868**, 13, 137-203.

[49] Paul, W.; Steinwedel, H., *US Pat.* 2939952, **1960**.

[50] March, R. E. *J. Mass Spectrom.* **1997**, 32, 351-369.

[51] Doroshenko, V.; Cotter, R. J. *J. Mass Spectrom.* **1997**, 31, 602-615.

[52] Yoshinari, K. *Rapid Commun. Mass Spectrom.* **2000**, 14, 215-223.

[53] Nourse, B. D.; Cooks, R. G. *Anal. Chim. Acta* **1990**, 228, 1-21.

[54] Splendore, M.; Lausevic, M.; Lausevic, Z.; March, R. E. *Rapid Commun. Mass Spectrom.* **1997**, 11, 228-233.

[55] Wilm, M., Mann, M. *Anal. Chem.* **1996**, 68, 1-8.

[56] Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.* **1995**, 67, 3802-3805.

[57] Fong K.W.Y.; Chan T.W.D. *J. Am. Soc. Mass Spectrom.* **1999**, 10, 72-75.

CHAPTER 2

Nanoelectrospray Techniques and Applications

2.1 Introduction

Since the introduction of the electrospray ionization (ESI) source for biomolecular analysis, considerable effort has been taken to achieve high sensitivity for peptide and protein analysis. Early work examined micro ESI needles fabricated from fused silica capillaries with inside diameters ranging from 5 to 250 μm [1]. The goal was to reduce droplet diameters and make the ESI process more efficient. A study of ion formation during ESI relating flow rate to droplet size, predicts droplet size is proportional to the two-thirds power of flow rate [2].

A desirable property of small droplets is a high surface-to-volume ratio. The smaller the droplet size, the higher the proportion of analyte available for desorption [3]. In order to obtain smaller droplets, one could operate at very low flow rates. Nanoelectrospray offers stable flow rates less than 100 nL/min, in addition to simplifying the analysis by not using pumps and by eliminating sample carryover by using a new needle for each analysis. Moreover, the small droplet size enables spraying of aqueous solutions and has a higher buffer tolerance than traditional electrospray [4].

A number of methods have been developed to establish electrical contact with the sample. One method involves inserting a gold plated tungsten or platinum wire directly into the nanoelectrospray capillary [5]. A nanoelectrospray is then established by applying a voltage to the metal strip. This method is simple and allows the user an unobstructed view of the contents of the glass capillary. A second approach involves

using vapour deposition techniques to apply a metal coating onto the glass of the electrospray emitter [2,6,7]. The electric field produced at the tip of these emitters is extremely high; as a result, the tips are susceptible to damage through electrical discharges. Several methods have been developed to increase the durability of these gold coatings. These methods involve the addition of organic layers prior to metal deposition [8] and also electroplating a second gold layer to reinforce the original metal layer [9].

The most frequently used method of depositing metal onto the glass surface involves vapor deposition of the chosen metal. There has, however, been a report of a so-called "fairy dust" technique, in which 2 μm gold particles are glued onto the glass which had been previously coated with a thin polyimide film [10]. The author tried the "fairy dust" technique; however, it was extremely difficult to apply the coating without damaging the tip of the needle.

This chapter will focus on manufacturing nanoelectrospray tips from aluminosilicate glass capillaries pulled using a commercial glass puller. The tips are manufactured in large batches and their properties will be examined with several examples.

2.2 Experimental

2.2.1 Materials

Cytochrome c (horse heart) and trypsin (bovine pancreas) were from Sigma Aldrich Canada (Oakville, ON). HPLC grade methanol, acetonitrile and acetic acid were from Fisher Scientific Canada (Edmonton, AB). Water used in these experiments

was from a Milli-Q Plus purification system (Millipore Corporation, Bedford, MA, USA).

2.2.2 Needle fabrication

Nanoelectrospray tips were pulled from aluminosilicate glass (1.2 mm o.d., 0.85 mm i.d. from Sutter Instrument Co.) with a P-97 Flaming/Brown micropipette puller. The starting material consisted of 10 cm long aluminosilicate glass, from which four tips could be produced.

An optical microscope allowed for a comparison of the fabricated tip to a commercial variety provided by Bruker Daltonics. A number of programs were evaluated in order to produce tips with acceptable taper lengths. The program chosen for manufacturing the tips contained the following parameters, Heat 445, Pull strength 50, Velocity 110, Cooling time 150.

In order to provide electrical contact with the capillary, there are two methods that have been developed. The glass can either be coated with a conductive layer, or a strip of metal can be inserted directly into the solution. The instrument used for collecting nanospray spectra was a commercial variety that used metal-coated capillaries. As a result, electrical contact through a conductive layer on the glass was required for any capillaries manufactured in house.

In order to hold the pulled capillaries during the metal deposition process, Teflon holders (10 x 10 x 1 cm) were manufactured. Each holder can accommodate 81 tips (9 x 9 grid, 1 cm on centre). In order to maintain a clean surface, the needles were

handled with tweezers when placing them vertically into the Teflon holders. The holes were drilled slightly smaller than the o.d. of the glass in order to provide a snug fit.

Metal deposition was done at the University of Alberta Microelectronic Fabrication Centre with a Kurt. J. Lesker sputter system. The device can hold four Teflon holders, resulting in 324 pieces of glass being sputtered at once. A 150 Å layer of titanium was first deposited onto the glass, followed by 1500 Å of gold. The coating process required approximately 25 min; however, the entire process required approximately 3 hours due to the time required to reach sufficient vacuum.

2.2.3 Cytochrome c Tryptic Digest

A solution of 38 μM equine cytochrome c in 48 mM NH_4HCO_3 was mixed with 0.5 μg trypsin and incubated for 45 minutes at 37 °C. An Eppendorf pipettor with a gel loader tip was used to transfer 5 μL of the digest into the nanoelectrospray emitter. In cases where small bubbles were observed near the tapered end, the capillary was briefly centrifuged and then placed into the commercial nanoelectrospray holder. An Esquire-LC ion trap (Bruker Daltonics, Billerica, MA) was used for production of nanoelectrospray spectra.

2.3 Results and Discussion

2.3.1 Background Species

The performance of the emitters was first measured by spraying the calibration mixture provided by Bruker Daltonics. This mixture contained a number of proprietary peptides and a MS^2 spectrum of one of the components is shown in Figure 2-1. In

addition, a number of acetonitrile blanks were run. A surprising discovery was the presence of several very strong peaks in the region from 200 to 700 m/z. The origin of these peaks is not clear; however, results show the contamination is present on the glass. Uncoated glass was rinsed with acetonitrile and sprayed with commercial emitters that were previously shown to be free of peaks in the m/z range under consideration. The identical set of background peaks was also observed.

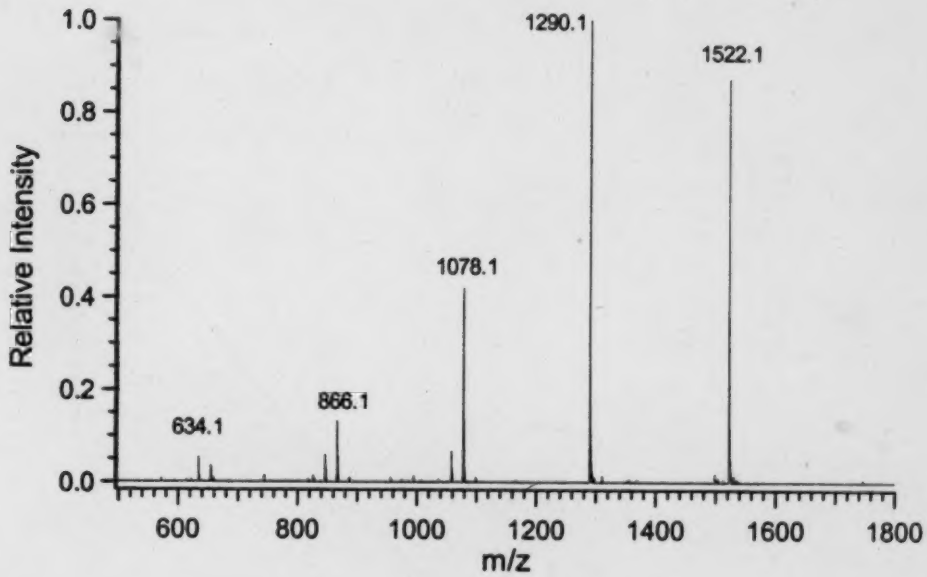

Figure 2-1 Nanoelectrospray MS^2 spectrum of 1522.1 m/z component of the Hewlett Packard ES tuning mix for Esquire-LC standard mass range.

Baking the tips to 218 °C overnight was enough to reduce the extent of contamination. Figure 2-2 (A) shows the background species present when spraying acetonitrile in emitters that were not subjected to cleaning steps. Figure 2-2 (B) shows the peaks present after baking the tips. The intensity of the peaks has been greatly reduced and the baked tips are more suitable for routine use.

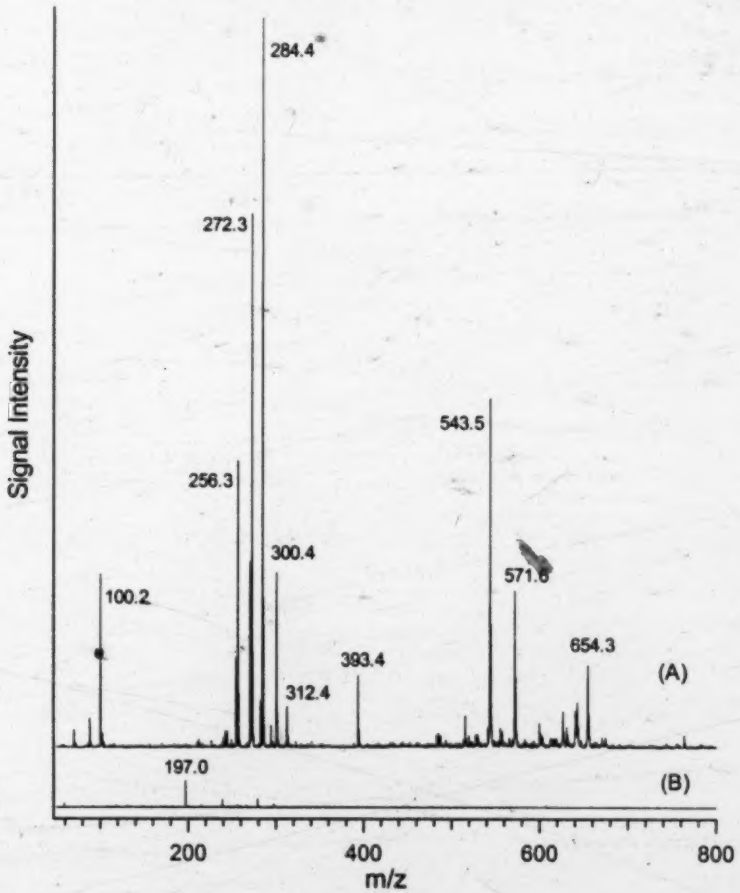

Figure 2-2 (A) Background species present when spraying acetonitrile in the nanoelectrospray emitters fabricated from aluminosilicate glass.
(B) Background species present after baking the emitters at 218 °C for 20 hours.

Two peaks from Figure 2-2 were chosen to fragment by MS/MS. Figure 2-3 shows the fragmentation pattern observed when the 571.6 Da ion was chosen for isolation and fragmentation. The two fragment masses at 300.4 and 272.3 Da appear strongly in the complete scan of Figure 2-2. Subsequent fragmentation of the 300.4 daughter ion also produces fragments present in Figure 2-2.

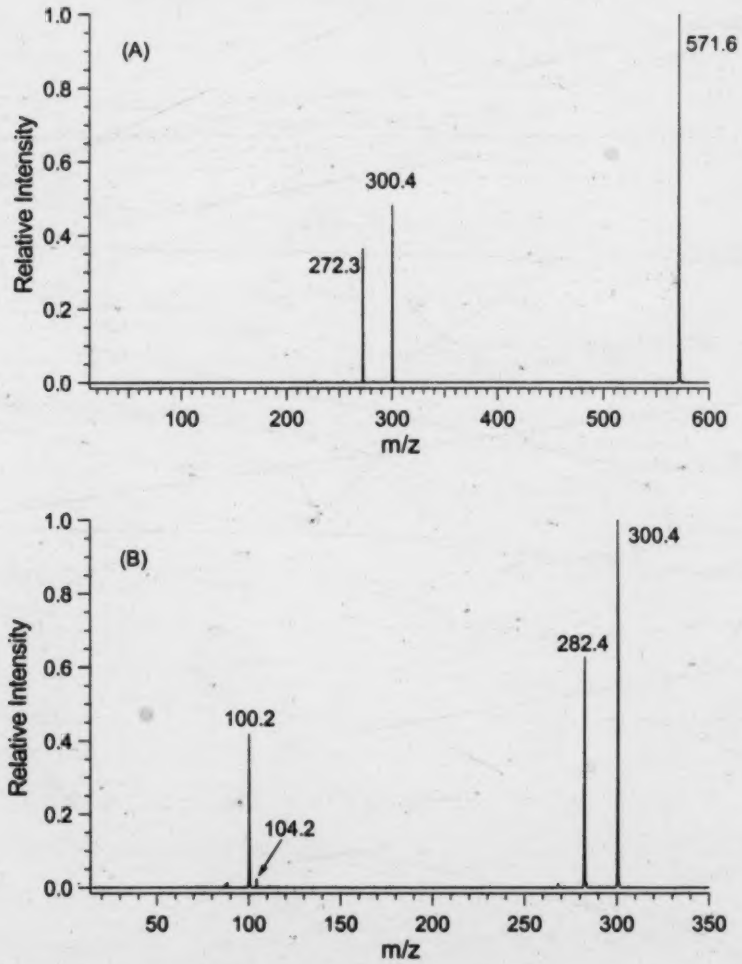

Figure 2-3 (A) MS^2 spectrum of the 571.6 ion from Figure 2-1
(B) MS^3 spectrum of the 300.4 ion from (A)

An identical examination of the peak at 543.5 was performed and the results of the fragmentation are shown below in Figure 2-4.

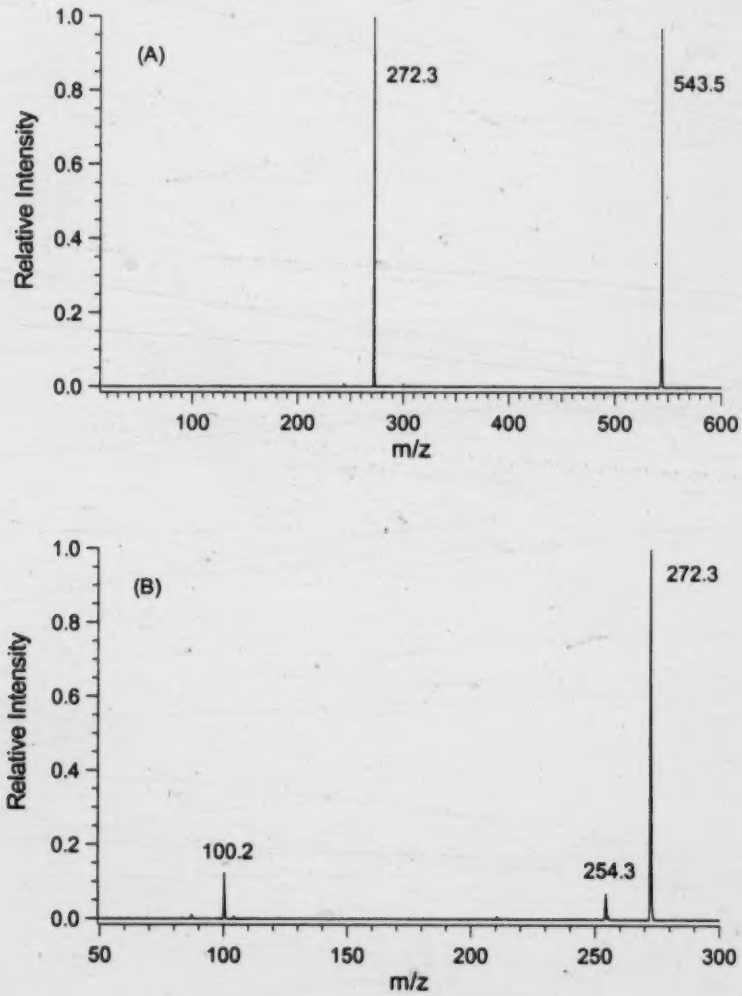

Figure 2-4 (A) MS^2 spectrum of the 543.5 ion from Figure 2-2
(B) MS^3 spectrum of the 272.3 ion from (A)

From the MS^n spectra produced on the background species, it is evident that the large number of the low mass peaks present in Figure 2-2 are due to fragmentation of the higher mass ions. However, the origins and identities of these species are unknown.

2.3.2 Imaging Nanoelectrospray Emitter Tips

Emitter tips were examined with an electron microscope at the University of Alberta Microelectronic Fabrication Center. Tips were sputtered in batches of 6 - 10 with a thin layer of gold prior to imaging. The high magnification allowed for an estimation of tip diameter, of which the majority were 5 μm when produced using the parameters mentioned in section 2.2.2. The smallest tip imaged had an inner diameter from about 1 – 2 μm. Figure 2-5 shows a side view of a nanoelectrospray emitter with a slightly larger diameter tip.

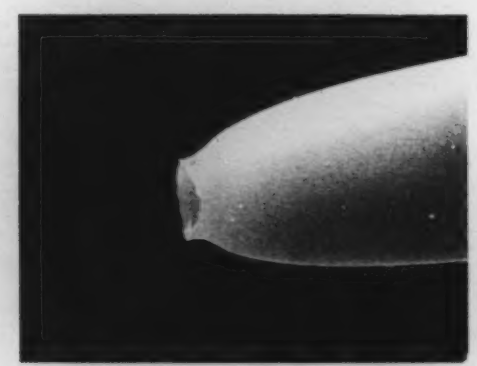

Figure 2-5 Nanoelectrospray emitter pulled from aluminosilicate glass from Sutter Instrument Co. (1.2 mm o.d., 0.85 mm i.d., catalog # AF120-85-10). Tip was pulled with a P-97 Flaming/Brown micropipette puller with program described in section 2.2.2.

The emitters functioned well as prepared and no attempt was made to determine the reproducibility of the manufacturing process.

2.3.3 Cytochrome c Tryptic Digest

Analyzing a solution of intact cytochrome c, as well as a tryptic digest, tested the performance of the fabricated nanoelectrospray emitters. Figure 2-6 shows the multiple charging effect obtained when spraying protein solutions by electrospray ionization.

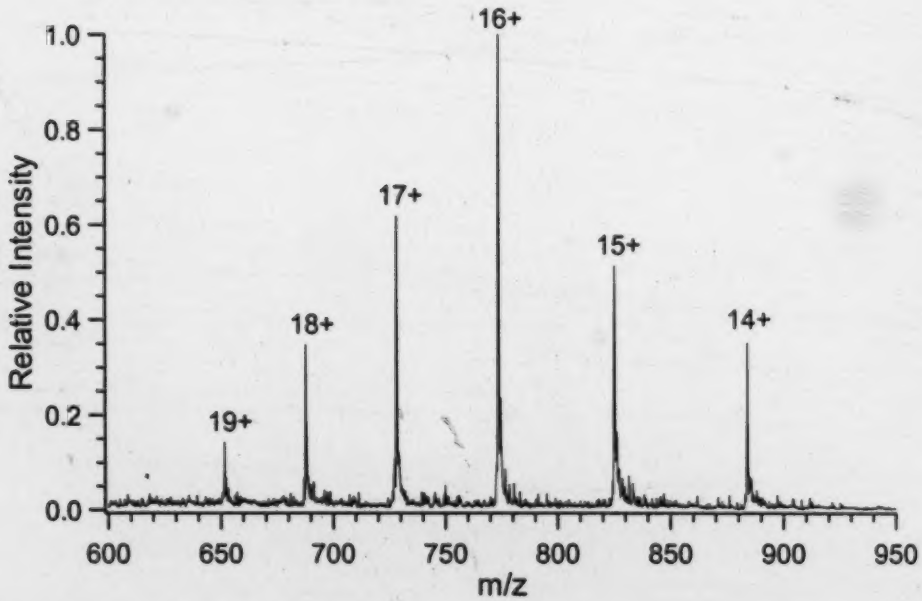

Figure 2-6 Nanoelectrospray spectrum of a 1 μM equine cytochrome c solution dissolved in water/methanol/acetic acid at 50/48/2 (v/v/v).

Figure 2-7 shows the peptide map obtained when spraying the tryptic digest of equine cytochrome c.

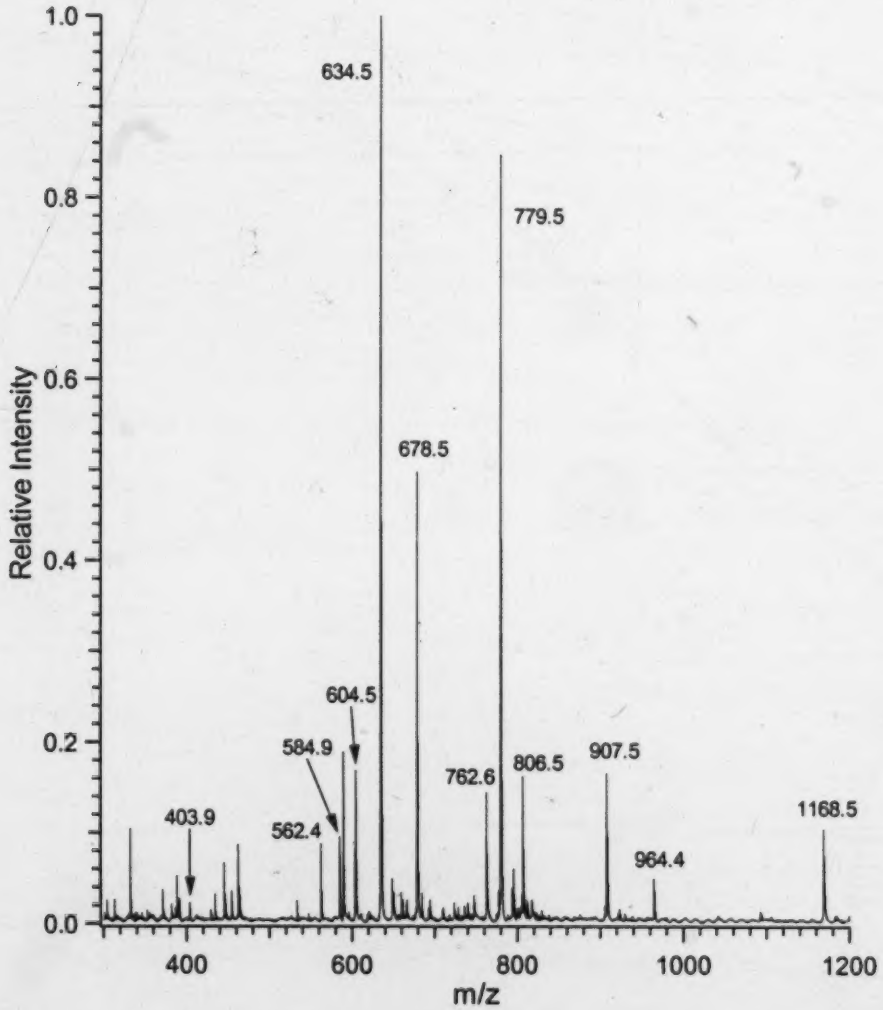

Figure 2-7 Tryptic peptide map of equine cytochrome c produced using the fabricated nanoelectrospray emitters. Labeled masses correspond to cytochrome c peptides. Final protein concentration was 19 μM dissolved in water/methanol/acetic acid to 50/49/1 (v/v/v).

The peptide map obtained from the nanoelectrospray analysis exhibited 81% coverage for equine cytochrome c. In addition, numerous peptides were chosen to fragment in order to obtain sequence information. Figure 2-8 shows the fragmentation pattern of the doubly charged peptide at 584.9 Da.

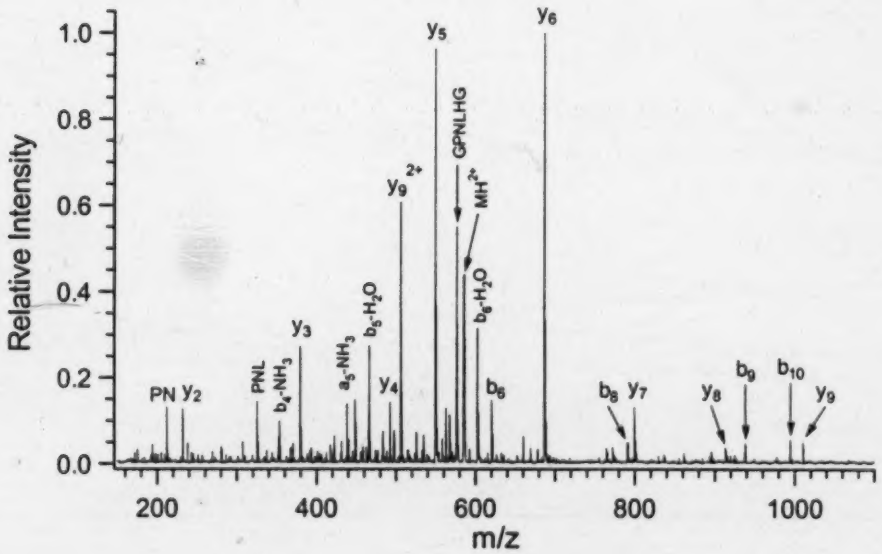

Figure 2-8 MS^2 of 584.9 peak from Figure 2-7 The doubly charged $M+2H^+$ species produces many singly charged products at higher m/z ratios. The fragments correspond to an amino acid sequence of TGPNLHGLFGR.

A second example of peptide fragmentation pattern is shown in Figure 2-9. This example shows the spectrum produced when fragmenting a singly charged peptide.

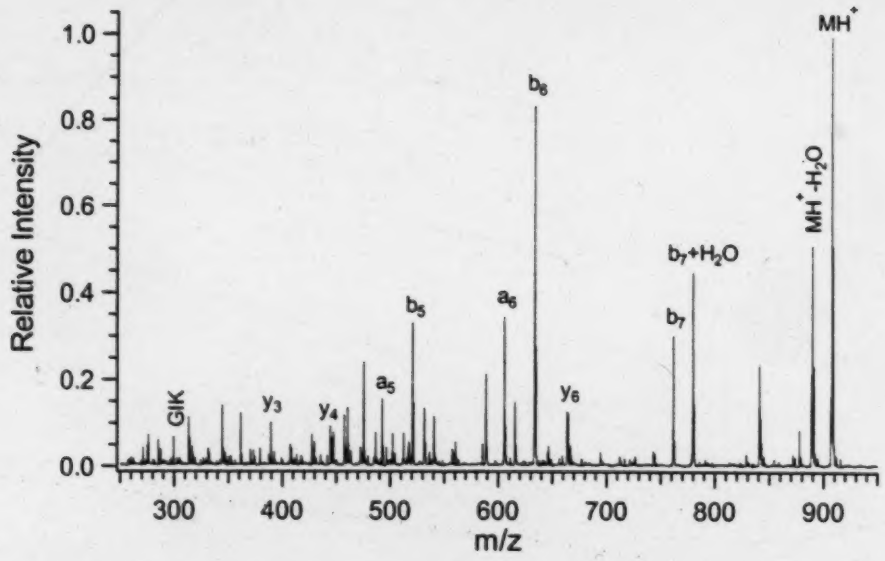

Figure 2-9 MS^2 of 907.5 peak from Figure 2-7. The fragments correspond to an amino acid sequence of MIFAGIKK.

2.3.4 Fragmentation of Highly Charged Bacterial Protein

The nanospray needles prepared using the method described in this chapter are being routinely used in our laboratory for a variety of applications. As an example, direct fragmentation of large protein ions generated by nanospray is studied. This work is motivated by the desire to produce fragment ion spectra of proteins for unique identification.

Further information for the proposed protein mass database (see Chapter 3) could consist of several of the more intense masses produced from fragmenting the proteins that are already catalogued. In order to obtain a fragmentation pattern of an *E. coli* protein, an HPLC fraction from a separation of *E. coli* lysate was chosen for nanoelectrospray analysis. Acetic acid was added at a concentration of 2% (v/v). Figure

2-10 (A) shows the mass spectrum obtained during the LC/ESI-MS analysis of the *E. coli* extract. Figure 2-10 (B) shows the charge envelope produced from nanoelectrospray MS. In order to study the potential of using MS^2 as a tool for producing MS fingerprint spectra, the +11 ion in Figure 2-10 (B) was chosen for isolation and fragmentation inside the ion trap.

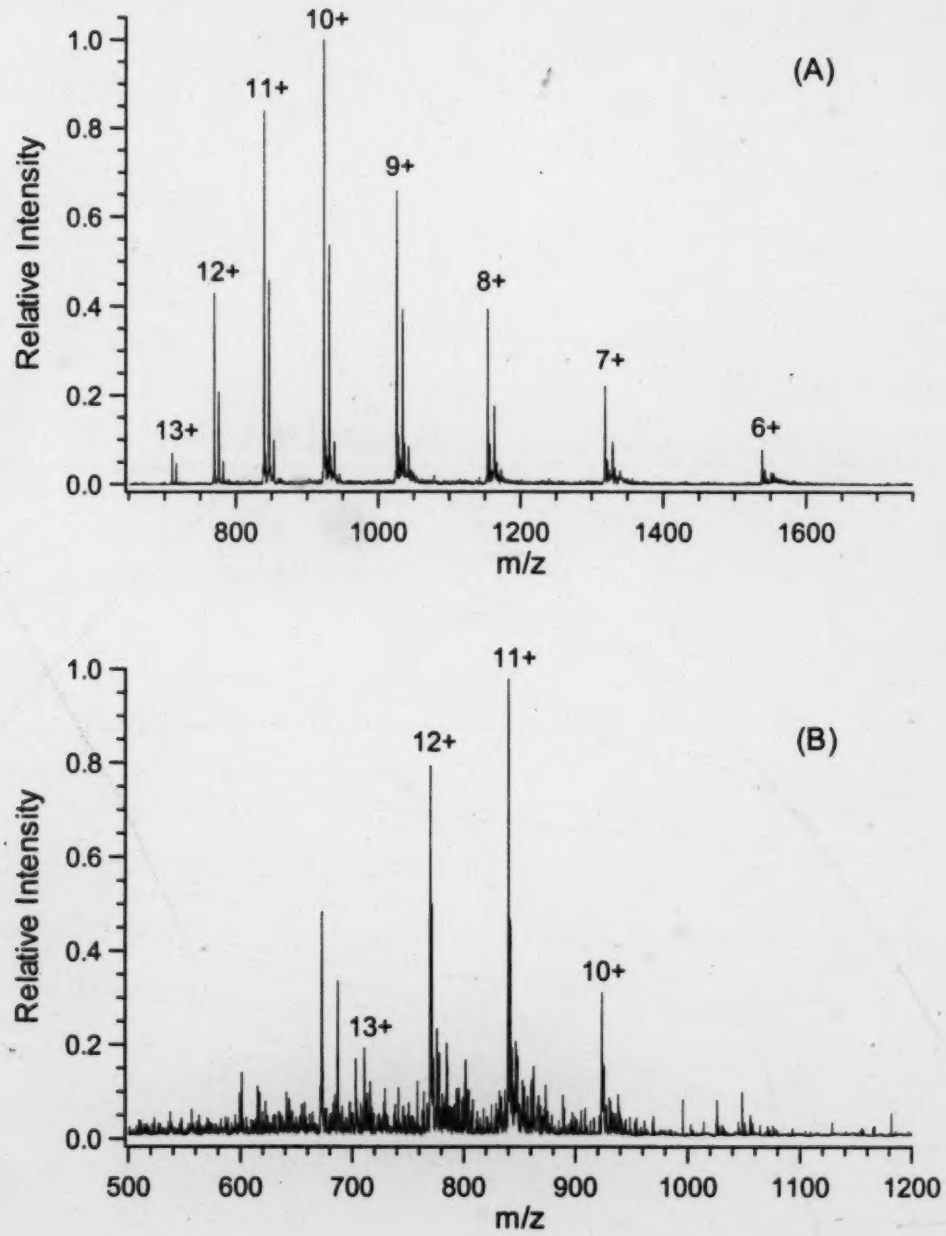

Figure 2-10 (A) Principal component in fraction 60 obtained from LC/ESI-MS $(M+H)^+ = 9226.0$ Da

(B) Fraction 60 analyzed by nanoelectrospray MS.

The spectrum produced after applying an excitation energy to the +11 charge state, is shown in Figure 2-11.

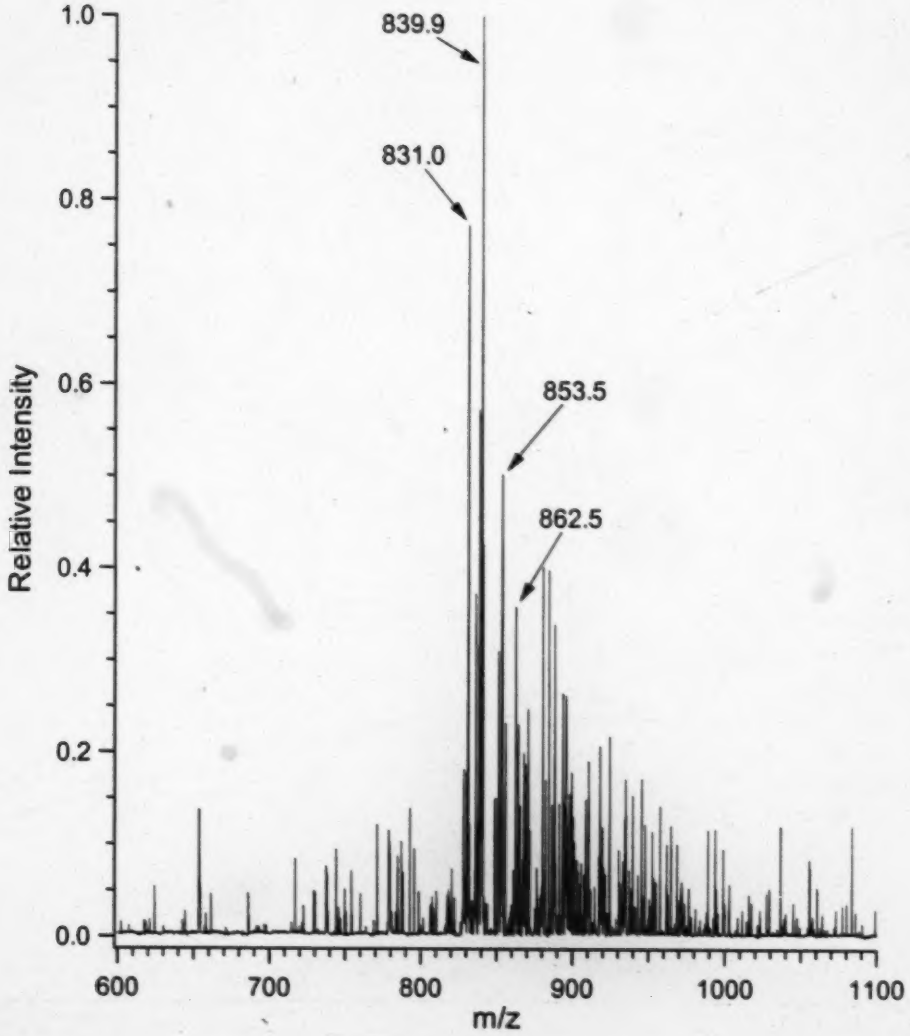

Figure 2-11 MS/MS spectrum of the +11 charge state (m/z = 839.9) Figure 2-10 (B).

The fragmentation pattern observed in Figure 2-11 is very difficult to interpret. Similar experiments were performed on multiply charged ions of cytochrome c with triple quadrupole instruments [11,12]. The relatively CID products have been attributed to the large internal energy required for dissociation of such large biomolecules. One study, however, has used the CID technique to differentiate 9 species of cytochrome c [13]. It therefore remains a possibility that a second dimension, consisting of characteristic masses observed during MS^2, may be a useful addition to the proposed mass database for bacterial identification.

2.4 Conclusions

Nanoelectrospray emitters were fabricated using a micropipette puller and gold sputtered in large batches. With up to 324 being fabricated at once, this method of producing nanoelectrospray emitters is a more economical solution to using commercial varieties, which are available for approximately $5 CAN each. Scanning electron microscope images revealed inner tip diameters of roughly 5 μm. The flow rates achieved using the fabricated emitters varied widely depending on which tip was chosen and the applied voltage; however, most tips resulted in flows of between 50 and 100 nL/min. With typical sample sizes of at least 2-3 μL, it is not uncommon for analysis times to extend past 30 minutes, which allows sufficient time for $MS^{(n)}$ sample characterization.

The properties of the fabricated nanoelectrospray emitters were tested on a tryptic digest of cytochrome c, as well as producing an MS^2 spectrum of an isolated fraction from an HPLC separation of *E. coli* ATCC 9637.

2.5 Literature Cited

[1] Emmet, M. R.; Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **1994**, 5, 605-613.

[2] Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, 136, 167-180.

[3] Wilm, M. S.; Mann, M. *Anal. Chem.* **1996**, 68, 1-8.

[4] Kuhnle, G. G.; Haferburg, D.; Grunow, M.; Hirsch, D.; Hahn, U. *Rapid Commun. Mass Spectrom.* **2000**, 14, 1307-1308.

[5] Fong, K. W.; Chan, D. *J. Am. Soc. Mass Spectrom.* **1999**, 10, 72-75.

[6] Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes*, **1994**, 136, 167-180.

[7] Korner, R.; Wilm, M.; Morand, K.; Schubert, M. *J. Am. Soc. Mass Spectrom.*, **1996**, 7, 150.

[8] Kriger, M. S.; Cook, K. D. *Anal. Chem.* **1995**, 67, 385-389.

[9] Bateman, K. P.; White, R. L.; Thibault, P. *Rapid Commun. Mass Spectrom.* **1997**, 11, 307-315.

[10] Barnidge, D. R.; Nilsson, S.; Markides, K. E. *Anal. Chem.* **1999**, 71, 4115-4118.

[11] Smith, R. D.; Barinaga, C. J. *Rapid Commun. Mass Spectrom.* **1990**, 4, 54-57.

[12] Smith, R. D.; Barinaga, C. J.; Udseth, H. R. *J. Phys. Chem.* **1989**, 93, 5019-5022.

[13] Smith, R. D.; Loo, J. A.; Barinaga, C. J.; Edmonds, C. G.; Udseth, H. R. *J. Am. Soc. Mass Spectrom.* **1989**, 1, 53-65.

CHAPTER 3

Detection of Low-mass Bacterial Proteins by LC/ESI-MS and MALDI* Mass Spectrometry

3.1 Introduction

Mass spectrometry (MS) is rapidly becoming a valuable tool for fast and accurate identification of microorganisms. One of the MS approaches being developed is based on the use of peptide and protein biomarkers [1]. A unique characteristic such as protein mass, or a combination of multiple properties such as protein mass and fragmentation pattern, can potentially be used as comparative factors for database creation and subsequent database searching.

Current MS techniques for analysis of bacteria have focused mainly on matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis of either intact bacterial cells [2-8] or cell lysates [9-16]. In addition, a few reports employing liquid chromatography/electrospray ionization-mass spectrometry (LC/ESI-MS) have surfaced [17-19].

This chapter will focus on the discussion of a combination of three techniques for the detection of low-mass (2–20 kDa) bacterial proteins from three bacteria, namely *Escherichia coli* (*E. coli*), *Bacillus megaterium* (*B. megaterium*) and *Citrobacter freundii* (*C. freundii*).

---

* MALDI of HPLC fractionations were obtained from Zhengping Wang

First, raw bacterial extracts are analyzed by MALDI without a separation step. In an effort to reduce suppression effects resulting from MALDI analysis of a complex combination of analytes, high-pressure liquid chromatography (HPLC) separation and fractionation, followed by MALDI analysis of the fractions will be presented. Finally, LC/ESI-MS of the bacterial lysates will be performed.

The simplest database for bacterial identification will likely consist of mass tables reflecting the protein components consistently detected from a mass spectrometric technique. This chapter addresses several technical and fundamental issues relevant to the detection of bacterial proteins for the purpose of database creation.

The number of proteins detected is contingent upon growth conditions as well as the cell wall structure. The more robust cell wall structure of gram-positive bacteria is more difficult to rupture than the thinner gram-negative walls. Although solubilization of cellular proteins from gram-positive bacteria has been facilitated through chemical lysis [10], other methods such as sonication [20], and French pressure cells [21] have been used to successfully rupture the more rigid cell walls. For unequivocal identification of bacteria from a mass database, it is necessary to examine as many proteins as possible in order to produce a set of data that reflects the unique set of low-mass proteins. The current public database is severely biased toward bacteria containing the greatest number of entries (e.g. *E. coli* and *B. subtilis*). Searching the online SWISS-PROT and TrEMBL databases with the Sequence Retrieval System [22] shows *E. coli*, with 2622 entries, as having the greatest number of proteins between 2 and 20 kDa. The next most abundant in the 2 to 20 kDa mass range is *B. subtilis* with 1496, followed by *H. pylori* with 1216.

3.2 Experimental

3.2.1 Materials

E. coli American Type Culture Collection (ATCC) 9637, *B. megaterium* and *C. freundii* were cultured at the Edgewood RDE Center at Aberdeen Proving Ground, MD. Spectrophotometric grade trifluoroacetic acid (TFA) was purchased from Sigma Aldrich Canada (Oakville, ON). HPLC grade acetonitrile and glacial acetic acid were from Fisher Scientific Canada (Edmonton, AB). Water was obtained from a Milli-Q Plus purification system (Millipore Corporation, Bedford, MA, USA). The MALDI matrix, α-cyano-4-hydroxycinnamic acid (HCCA) was purchased from Sigma Aldrich Canada (Oakville, ON).

3.2.2 Sample Preparation

Crude water-soluble protein extracts were obtained through a solvent suspension method. About 25 mg of lyophilized *E. coli* ATCC 9637, *B. megaterium* and *C. freundii* was suspended in 1 ml 0.1% (v/v) aqueous TFA, vortexed for 3-5 minutes and centrifuged. The supernatants were removed and this procedure was repeated until no apparent protein signals were detected by MALDI. A total of 5 to 6 extractions were performed, producing a total volume of 5 to 6 ml. The supernatants from each of the 3 bacterial samples were combined and filtered with Microcon 3000 Da molecular weight cut-off filters (Amicon, Oakville, Ontario). The three cell extracts were then concentrated to 500 μL by SpeedVac.

3.2.3 Instrumentation

High Performance Liquid Chromatography

Solvent delivery and separations were performed on an Agilent (Palo Alto, CA, USA) 1100 series HPLC equipped with an autosampler. All connections were made with 0.127 mm i.d. PEEK tubing and fingertight fittings (Upchurch Scientific). Separations were optimized for each bacterial extract and the conditions are listed in Table 3-1. The LC separations were made using a reversed phase Vydac C_8 column (2.1 mm i.d. x 150 mm; particle size, 5 μm; pore size, 300Å). The LC mobile phase consisted of water with 0.1% TFA (v/v) and acetonitrile with 0.1% TFA (v/v). A flow rate of 200 μl/min was employed for all samples used during the LC/ESI-MS analysis, with a sample injection volume of 40 μL.

For LC off-line MALDI, 100 μL of bacterial extract was separated on a reversed phase Vydac C_8 column (4.6 mm i.d. x 250 mm; particle size, 5 μm; pore size, 300Å) at a flow rate of 500 μl/min. The following gradient was used for the separation (time:%acetonitrile) 0:2, 10:20, 50:40, 60:55.

The mobile phase was the same as that employed for the online LC/ESI-MS. Every minute throughout the separation, a Gilson FC 203B fractionation collector collected fractions of one-minute duration.

Table 3-1 Optimized separation conditions (shown as % ACN)

E. coli	*B. megaterium*	*C. freundii*
0 min – 2%	0 min – 2%	0 min – 2%
10 min – 20%	10 min – 20%	10 min – 20%
40 min – 40%	40 min – 40%	30 min – 40%
45 min – 55%	45 min – 55%	60 min – 55%
60 min – 90%	60 min – 90%	

Electrospray Ionization Mass Spectrometry

The HPLC effluent was analyzed with an Agilent (Palo Alto, CA) 1100 series MSD quadrupole mass spectrometer. The mass range from 500 to 3000 m/z was scanned in 1.90 sec and the ions were detected with a high-energy dynode detector. HP Cessation software provided control of both the HPLC and MS. The Igor Pro software package (WaveMetrics, Inc., Lake Oswego, OR) was used to reprocess the acquired data.

Matrix-Assisted Laser Desorption Ionization Mass Spectrometry

The MALDI results were obtained on a time-lag focusing linear time-of-flight mass spectrometer with a 1 m flight tube [23]. A 337 nm nitrogen laser (model VSL 337ND, Laser Sciences Inc., Newton, MA) was used for desorption/ionization. In general 50-100 laser shots are averaged to produce a mass spectrum.

For direct MALDI analysis, about 1 mg lyophilized bacterial sample was extracted by 500 μL 0.1% (v/v) TFA. For off-line MALDI, the HPLC fractions were concentrated 50 times to about 10 μL before analysis. A two-layer sample preparation

method with HCCA as the matrix was used for MALDI analysis. In the two-layer method [24], the first layer is formed by applying 1 μL of 100 mM HCCA in acetone/water (99/1, v/v) to the MALDI probe tip and allowing it to dry very quickly in air. For the second layer, sample solution was mixed 1:1 with saturated HCCA in formic acid/isopropanol/water/ (1/2/3 v/v/v); about 0.6-1 μL of the second layer solution is then applied onto the first layer and allowed to dry. The sample is then washed with 1 μL of distilled water for 10 seconds and then dried with a stream of nitrogen.

3.3 Results and Discussion

3.3.1 LC/ESI-MS

Using the deconvolution software in HP ChemStation, the electrospray spectra were studied and deconvoluted to produce the protonated molecular weights. A large number of charged species increases the accuracy of the deconvoluted mass and it is expected to be about 0.02 %. Post-column addition of acetic acid not only increased the number of ions detected in the ion envelope, but also, increased the intensity of the individual peaks.

Due to the signal suppressing properties that mobile phases containing TFA have on the ESI process [25], post-column addition of glacial acetic acid was delivered at 100 μl/min. The acid was added to the column effluent using a poly(ether ether ketone) (PEEK) "Y" connector and a syringe pump (Cole Parmer). The "Y" was connected to the electrospray interface with 30 cm of PEEK tubing. The length of the

tubing exiting the "Y" and entering the ESI source had to be long enough to provide adequate mixing of the acid with the column effluent.

A large voltage drop in the region between the capillary exit and the first skimmer aided in ion transmission and produced more intense signals in the TIC; however, it also had the effect of stripping charges from the analytes as well as producing severe fragmentation due to collision induced dissociation (CID). In order to overcome the detrimental CID effects caused by a large and constant fragmentation voltage, the voltage was varied as the quadrupole scanned across the selected mass range. The optimized voltage ramp was found to be: m/z 500 = 60V, 1000 = 120V, 3000 = 220V. The resulting mass spectra showed an increased number of peaks and a higher signal to noise ratio versus those obtained using a constant fragmentation voltage. Figure 3-1 follows a specific protein's m/z envelope change as the fragmentation voltage is varied. The following parameters were used for MS detection of the chromatographic effluent: gain, 5; step size, 0.15; drying gas, 10 L/min; drying gas temperature, 350°C; nebulizer pressure, 25 psi; capillary voltage, 4500 V.

The optimized TIC's of *E. coli*, *B. megaterium* and *C. freundii* are shown in Figure 3-2. Note the unusual chromatographic peak shape. A more thorough discussion of this phenomena is discussed in Chapter 4. Briefly, removing the static mixer on the pump produced the observed peak shapes. The oscillating nature of the ion current can then used to help the peak identification process. Since peak shape itself is not used to obtain useful information, the shape of the eluting peaks is not important, rather it is vital to obtain partial separation in order to reduce the burden on the mass analyzer. Results for the three bacteria samples are shown in Tables 3-2 through 3-4.

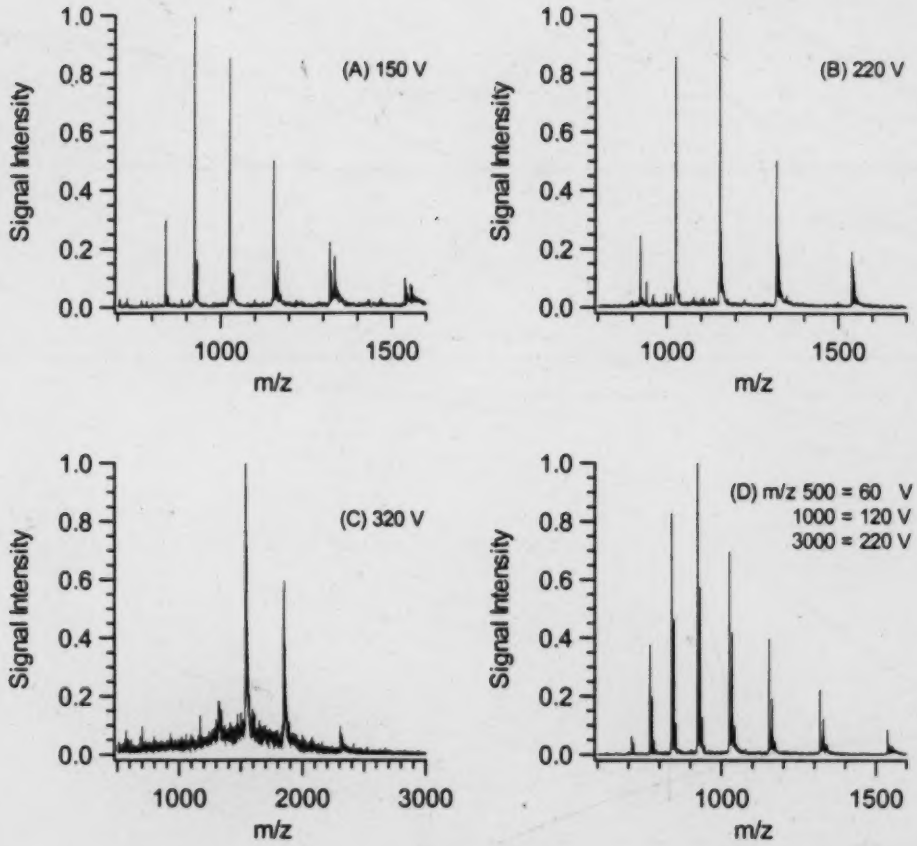

Figure 3-1 Effect of increasing fragmentation voltage on charge state envelope. A higher voltage drop in the CID region (see Figure 1-3 for instrument schematic) strips protons from the protein and results in a shift to lower charge states. (D) shows the wider charge envelope and increase in signal observed when the fragmentation voltage is ramped.

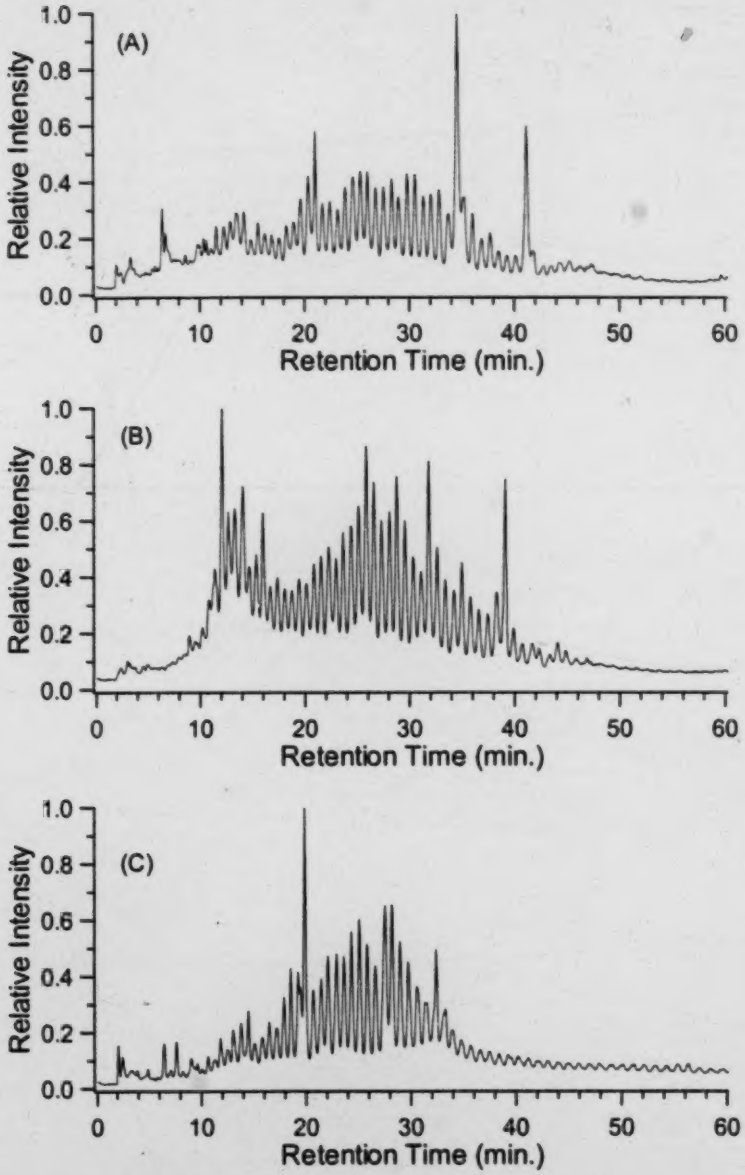

Figure 3-2 Total ion chromatograms observed with static mixer removed.

(A) *E. coli*

(B) *B. megaterium*

(C) *C. freundii*

Table 3-2 $(M+H)^+$ from *E. coli* extract by online LC/ESI-MS

2008.0	**2735.9**	5170.8	**7274.2**	**9226.4**	**9740.2**	**15693**
2123.0	**3509.6**	**5550.1**	**7707.3**	**9230.5**	**10386**	15767
2139.0	**3624.7**	**6254.9**	**7781.6**	**9264.5**	**10459**	
2375.3	**3793.1**	**6315.7**	**7855.4**	**9518.2**	11224	
2416.2	4480.8	6330.3	**9064.3**	**9536.3**	11297	
2431.5	**5036.7**	**7140.0**	**9191.7**	**9610.0**	**11782**	
2504.9	**5096.7**	**7272.1**	**9209.1**	9684.0	**13094**	

Table 3-3 $(M+H)^+$ from *B. megaterium* extract by online LC/ESI-MS

3047.5	**6335.7**	7157.4	**7710.5**	9754.0	**11063**
3186.9	6352.1	**7279.7**	9331.5	9767.8	**11538**
4379.2	6389.0	**7424.9**	**9335.0**	9828.5	**11612**
4741.1	6392.7	7451.4	**9349.2**	9884.1	**11726**
4816.3	**6448.7**	**7453.5**	**9352.4**	10026	12046
6261.9	**6577.6**	**7466.9**	**9620.1**	**10045**	12112
6275.7	**6897.3**	**7519.3**	9693.9	**10453**	12408
6315.5	7110.7	7648.9	**9747.4**	10696	

Table 3-4 $(M+H)^+$ from *C. freundii* extract by online LC/ESI-MS

2063.1	3038.5	**4504.2**	**7751.8**	**10107**	**11870**
2391.3	**3236.6**	**5240.3**	**8278.3**	**10301**	11956
2431.0	**3449.6**	**5253.6**	8547.5	**10682**	**12157**
2473.2	**3565.0**	**5934.7**	9057.0	11162	12239
2515.2	**3580.8**	6530.3	**9196.4**	11252	16745
2676.0	**4007.3**	**6720.7**	9223.5	11675	**17178**
2793.0	**4420.5**	**7334.6**	**9522.6**	11691	
2852.8	**4436.7**	**7736.0**	**9527.4**	**11750**	

* The bolded masses match those found using MALDI and the underlined masses are those that match the Swiss-Prot or TrEMBL databases.

3.3.2 MALDI

3.3.2.1 Direct MALDI analysis. Figure 3-3 shows the mass spectra of the extracts from *E. coli*, *B. megaterium*, and *C. freundii*. Tables 3-5 to 3-7 list the masses of singly charged molecular ions obtained from direct MALDI analysis. With external calibration, the mass measurement accuracy is expected to be about 0.05%. Figure 3-4 shows the direct MALDI spectra obtained when the volume of extraction solvent is varied.

In direct MALDI analysis, a small extraction volume produces spectra in which only a few peaks dominate. Figure 3-4 compares the effect of extracting the same amount of bacteria into 200 μL and 1500 μL. The number of peaks increase drastically once the volume of the extraction solvent is increased. One possible conclusion could be due to the different solubilities of low mass proteins in the bacteria. In a small volume, some highly soluble components tend to saturate the extraction solvent, especially when their abundance is high. As the volume of the extraction solvent is increased, more components are efficiently extracted. Therefore, the ratio of the amount of lyophilized sample and the volume of extraction solvent should be optimized in order to effectively extract more components and detect more peaks in the direct MALDI analysis. Additionally, the greater suppression that exists in the more concentrated crude mixture may be another reason why only a few peaks dominate in the direct MALDI analysis.

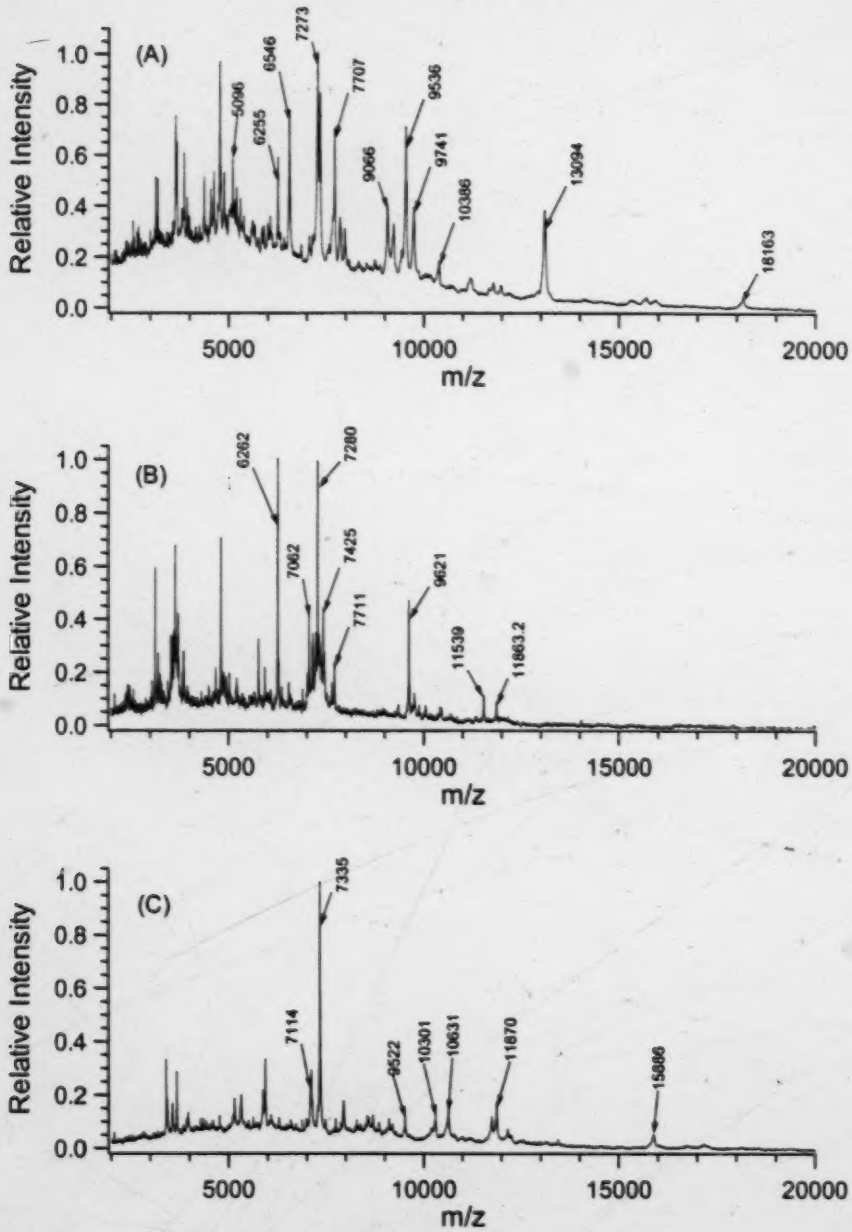

Figure 3-3 Direct MALDI spectra of (A) *E. coli*, (B) *B. megaterium* and (C) *C. freundii* crude extract

Table 3-5 $(M+H)^+$ from *E. coli* crude mixture by direct MALDI

2125.5	5381.5	6412.7	7707.3	9535.6	11977
2374.3	5993.2	6852.7	7849.1	9738.5	13095
2384.4	6056.9	7061.4	7969.7	10385	15682
5096.6	6255.1	7273.0	9066.0	11206	18164
5293.3	6316.3	7333.2	9226.0	11783	

Table 3-6 $(M+H)^+$ from *B. megaterium* crude extract by direct MALDI

6261.1	7279.8	9350.8	10454
6540.7	7425.2	9621.5	11538
6893.2	7449.7	9756.3	11863
7061.1	7647.9	9881.7	
7140.0	7711.0	10044	
7157.7	9335.9	10418	

Table 3-7 $(M+H)^+$ from *C. freundii* crude extract by direct MALDI

7089.9	7736.0	8695.1	10287	11871
7114.6	7865.3	8850.6	10301	12156
7141.3	8278.0	9106.9	10631	13448
7335.4	8546.9	9523.2	11751	15887

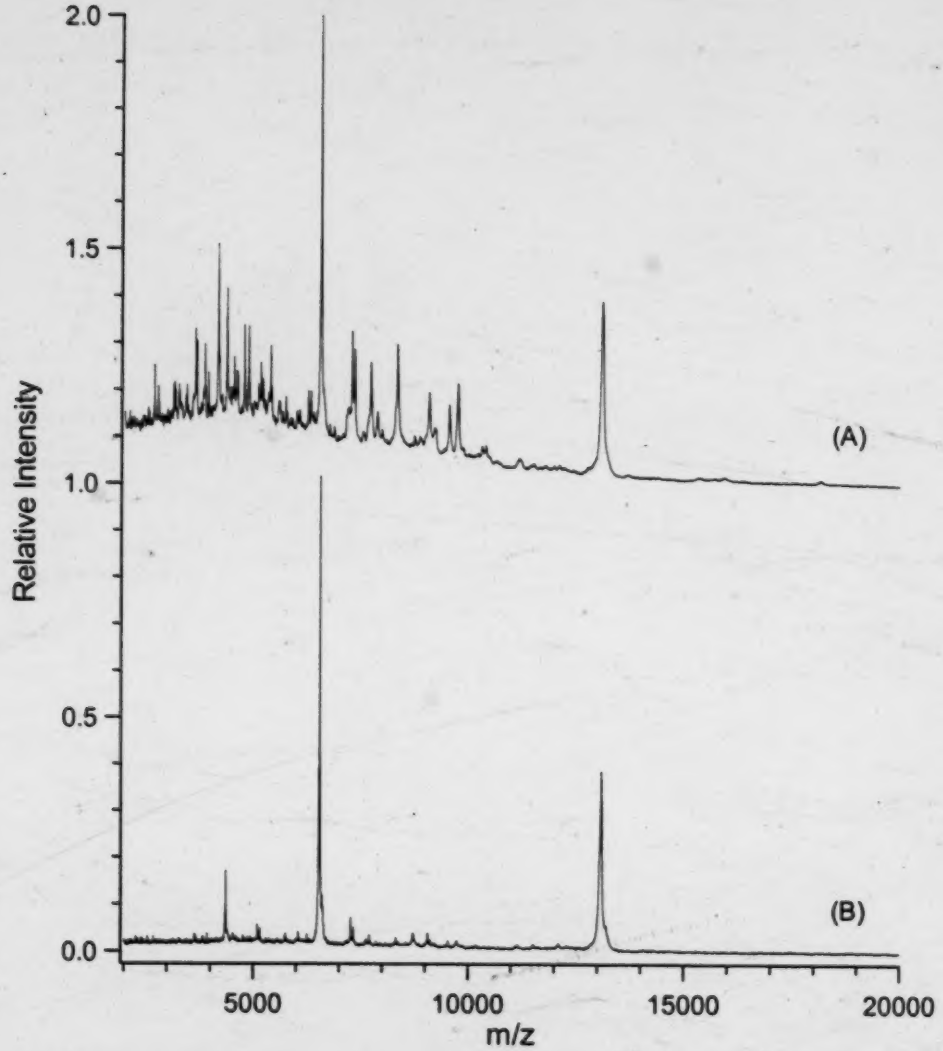

Figure 3-4 Direct MALDI spectra of *E. coli* crude extract in different extraction volumes.

(A) 1.5 mg *E. coli* in 1.5 mL 0.1% (v/v) TFA

(B) 1.5 mg *E. coli* in 200 μL 0.1% (v/v) TFA

The *E. coli* used here is not from the same batch as that in Figure 3-3 (A).

3.3.2.2 LC/off-line MALDI The protein masses detected by LC/off-line MALDI are listed in Tables 3-8 through 3-10 (Note: bolded masses match LC/ESI-MS data and underlined match public database). The number of additional masses obtained through the offline MALDI method compared to the direct analysis is striking. It is clear that although MALDI is able to analyze mixtures of proteins, there is a limit and therefore a separation step is required to reduce suppression effects caused by too many analytes.

During MALDI analysis, the oxidized form of some proteins is often detected in addition to their non-oxidized form. The oxidation of methionine most likely occurred during MALDI sample preparation; for this reason, the mass tables contain only the $(M+H)^+$ of the non-oxidized forms.

Table 3-8 $(M+H)^+$ from *E. coli* extract by LC/offline MALDI*

2003.9	2426.4	2846.3	**3510.2**	4444.6	5644.3	6669.0	7734.5	8996.7	10176
2006.4	**2431.2**	2858.0	3534.0	4447.0	5649.9	6684.1	7740.2	9048.1	10257
2009.1	2453.3	2860.2	3538.1	4456.4	5658.9	6700.0	7751.5	9054.5	10301
2014.0	2463.0	2873.8	3546.8	4465.2	5666.5	6703.2	**7781.2**	**9066.2**	10315
2025.8	2474.0	2877.9	3557.1	4473.0	5694.8	6720.7	7798.7	9079.6	10332
2040.7	2487.1	2883.8	3586.6	4483.7	5709.3	6725.2	7852.1	9160.3	10372
2044.1	2506.8	2910.0	3592.3	4499.8	5719.8	6772.3	7868.5	9175.3	10378
2050.0	2512.7	2969.7	3599.4	4562.4	5725.4	6786.9	8188.5	**9192.2**	**10387**
2053.9	2526.4	2975.1	3615.4	4716.4	5736.1	6826.5	8205.9	**9209.1**	10437
2066.6	2529.8	2984.3	**3624.0**	4722.1	5746.7	6853.5	8214.3	**9226.5**	10453
2074.7	2544.5	2999.1	3647.7	4792.3	5753.9	6866.3	8218.3	**9228.0**	**10463**
2084.5	2559.9	3002.7	3666.6	4894.9	5772.2	6890.1	8220.4	9252.0	10477
2106.0	2562.9	3014.1	3749.6	4903.3	5800.2	6942.1	8227.8	**9264.3**	10618
2123.3	2574.0	3022.1	3764.1	4940.4	5808.2	6951.4	8242.2	9281.0	10653
2126.8	2577.1	3037.0	3769.3	4984.8	5818.0	6958.2	8257.7	9295.7	10660
2132.2	2588.2	3054.4	3782.1	5005.6	5850.2	7060.3	8266.7	9368.7	10664
2140.7	2599.2	3073.3	**3793.0**	5017.0	5856.9	7069.6	8280.2	9424.1	10945
2165.5	2602.2	3090.3	3795.8	**5035.2**	5867.9	7094.6	8290.8	9430.7	11035
2196.9	2608.3	3100.8	3907.7	5040.3	5872.5	7108.5	8305.3	9439.1	11170
2207.8	2612.0	3119.2	3930.2	5051.6	5901.1	**7138.5**	8324.6	9458.3	11184
2232.3	2620.7	3126.7	3954.0	5072.8	5994.2	7158.0	8341.5	9475.3	11208
2236.9	2629.8	3135.2	3983.8	5086.6	6011.7	7169.3	8369.4	9478.0	11216
2240.3	2634.6	3158.6	3996.0	**5096.7**	6139.6	7184.6	8375.6	**9519.6**	11240
2259.6	2639.7	3191.5	4002.9	5143.5	6179.6	7254.7	8397.7	9527.0	11472
2263.9	2649.1	3212.4	4007.7	5154.2	6196.3	7265.4	8449.5	**9536.5**	11653
2278.9	2656.0	3226.2	4011.6	5179.8	6223.6	7269.1	8524.8	9544.5	11779
2282.4	2660.8	3250.6	4023.6	5202.3	6243.3	**7273.7**	8590.6	9554.6	**11785**
2292.5	2664.8	3258.8	4034.4	5252.9	**6255.4**	**7275.4**	8634.8	9572.8	11794
2296.8	2668.3	3285.6	4053.3	5295.3	6261.1	7280.9	8670.4	9583.1	11870
2300.7	2675.1	3289.2	4068.9	5362.1	6265.3	7282.8	8782.0	**9610.8**	11980
2305.3	2680.6	3294.6	4086.2	5378.2	6282.9	7289.5	8796.0	**9739.8**	12233
2311.6	2683.3	3302.1	4094.6	5381.6	6297.8	7292.7	8800.0	9749.5	12446
2330.7	2693.1	3309.1	4104.0	5395.9	**6316.2**	7297.5	8813.9	9753.8	12769
2339.0	**2737.1**	3324.4	4111.5	5414.1	6323.6	7307.3	8820.4	9765.7	13077
2352.8	2744.8	3331.1	4127.9	5430.3	6338.0	7313.6	8858.6	9785.0	**13095**
2359.9	2747.7	3388.5	4145.0	5449.3	6344.4	7321.1	8868.2	9833.7	13109
2373.7	2754.2	3400.7	4251.1	5466.9	6368.9	7326.2	8877.2	9851.6	13127
2375.5	2779.7	3406.3	4260.3	5471.6	6411.1	7333.7	8880.8	9884.5	13240
2382.0	2785.9	3416.9	4365.6	5480.2	6453.3	7412.0	8885.4	9951.6	13650
2386.1	2795.4	3424.1	4371.8	5493.0	6485.5	7569.6	8892.4	9981.7	14749
2395.4	2808.7	3475.9	4384.1	5548.8	6493.3	7599.9	8896.9	9995.5	14839
2400.3	2813.9	3486.0	4393.6	**5551.1**	6555.2	7616.0	8964.8	10045	**15699**
2402.8	2819.3	3491.8	4399.0	5566.4	6600.1	7619.3	8978.4	10106	18162
2417.0	2842.5	3499.7	4406.9	5585.0	6616.5	**7707.8**	8993.6	10123	

* Bolded masses match LC/ESI-MS data and underlined match SwissProt or TrEMBL databases.

Table 3-9 $(M+H)^+$ from *B. megaterium* extract by LC/offline MALDI*

2022.5	2977.0	3542.6	4408.4	5015.8	5438.4	6362.3	7184.2	8720.0	10407
2166.4	2989.7	3554.7	4482.1	5029.8	5442.8	6382.4	7198.6	8793.6	10412
2210.7	2997.3	3628.9	4491.9	5026.8	5446.9	**6394.2**	**7279.8**	8817.7	10423
2222.5	3005.3	3636.5	4504.7	5040.2	5451.1	6414.3	7297.4	8849.0	10435
2232.4	3041.3	3693.6	4513.5	5055.2	5486.0	6445.7	7311.0	8861.7	**10453**
2248.1	3065.9	3700.6	4523.3	5059.3	5525.0	**6450.1**	7328.3	8866.2	10633
2270.4	3102.3	3751.8	4534.8	5069.6	5554.0	6491.9	7338.7	8942.0	10684
2281.1	3170.4	3796.0	4537.1	5073.2	5562.1	6497.5	7351.8	8979.7	**11063**
2289.6	**3186.8**	3800.0	4540.4	5089.9	5631.2	6508.0	7357.3	9188.5	11135
2295.9	3232.9	3811.0	4578.9	5094.5	5750.9	6562.5	**7425.4**	9328.3	11152
2313.0	3250.1	3817.0	4678.2	5105.5	5763.1	6566.0	7436.4	**9332.9**	11273
2325.1	3261.8	3822.1	4703.1	5112.5	5795.5	**6577.8**	**7453.1**	9348.2	11469
2353.8	3271.2	3839.1	4717.2	5139.1	5888.4	6580.2	**7467.1**	**9350.8**	11523
2363.8	3306.0	3878.9	4736.6	5143.1	5936.4	6591.2	7497.1	**9353.6**	**11539**
2730.5	3489.4	4235.9	4903.6	5271.9	6246.4	6966.5	8050.9	9731.6	11554
2743.7	3496.4	4304.2	4946.4	5291.8	**6261.9**	6973.0	8125.9	**9748.0**	11614
2420.9	3312.3	3935.3	**4742.1**	5147.1	5950.1	6595.2	7507.0	9483.5	**11728**
2435.1	3323.3	3940.6	4789.7	5164.5	5967.7	6690.3	**7520.2**	9602.4	12081
2524.1	3348.9	3977.9	4798.6	5185.6	6025.0	6757.0	7527.7	9015.8	12414
2538.3	3374.9	3995.9	4807.7	5191.3	6028.6	6769.5	7572.3	**9621.6**	12475
2614.1	3389.4	4027.0	**4816.5**	5203.1	6032.2	6771.1	7593.8	9625.4	14033
2649.3	3397.3	4041.6	4858.5	5208.5	6037.6	6843.8	7611.1	9635.6	14105
2656.1	3422.6	4056.5	4863.5	5222.4	6082.3	6851.4	7704.0	9654.6	14220
2709.0	3428.2	4119.3	4872.6	5235.0	6089.7	**6899.0**	**7711.0**	9658.0	14224
2722.6	3446.0	4233.9	4890.7	5240.7	6226.1	6933.3	7730.8	9697.6	15227
2801.6	3511.3	4309.6	4955.4	5296.5	6278.1	7061.9	8218.9	9909.6	
2904.2	3514.1	4316.8	4972.9	5342.3	6296.5	7108.1	8570.2	9983.3	
2839.2	3520.0	4319.1	4977.7	5372.6	**6335.1**	7117.3	8586.7	**10045**	
2945.9	3537.5	**4380.4**	4985.8	5395.1	6356.4	7170.6	8604.1	10061	

* Bolded masses match LC/ESI-MS data

Table 3-10 $(M+H)^+$ from *C. freundii* extract by LC/off-line MALDI*

2064.3	2662.0	**3237.1**	3971.4	**5240.9**	6547.1	8801.6	10948
2210.3	**2676.7**	3253.5	3991.2	**5255.5**	6551.9	8842.5	11416
2215.2	2688.6	3290.3	**4007.5**	5299.3	6666.0	**9196.6**	11424
2221.1	2744.2	3360.6	4036.0	5379.3	6684.8	9285.0	**11748**
2238.3	2755.0	3370.1	4068.8	5457.8	**6721.2**	9352.5	11761
2251.5	2769.7	**3448.9**	4208.4	5493.0	6774.0	9364.8	**11870**
2319.3	**2792.9**	3463.3	**4420.5**	5548.9	6791.6	9367.9	11902
2334.5	2806.6	3467.3	**4435.0**	5627.1	6942.5	**9521.9**	12069
2352.0	2830.9	3481.1	4453.7	5646.1	7022.5	**9527.5**	**12157**
2391.0	2848.6	3504.9	4475.6	5752.9	**7335.2**	9756.5	12866
2431.1	**2853.2**	3530.8	**4503.6**	5792.1	**7735.6**	9866.8	15519
2475.6	2878.9	3545.8	4567.5	5850.5	**7751.5**	**10108**	15903
2500.3	2957.5	**3564.3**	4681.0	5856.9	7865.5	10121	159120
2514.9	2981.5	**3581.4**	4762.0	**5935.6**	7880.9	10188	15940
2519.9	3028.8	3669.4	4977.2	5995.0	8226.0	**10300**	16774
2579.2	3059.7	3747.7	4997.9	6118.7	**8278.2**	10447	**17177**
2583.9	3064.5	3922.5	5090.3	6184.7	8318.0	10522	18552
2621.0	3138.4	3931.8	5117.5	6205.4	8487.1	10632	
2636.1	3175.5	3950.9	5171.2	6255.5	8698.5	10649	
2646.4	3188.1	3955.0	5188.0	6299.5	8782.3	**10681**	

* Bolded masses match LC/ESI-MS data

3.3.3 Method Evaluation

For *E. coli*, *B. megaterium* and *C. freundii*, direct MALDI analysis shows 29, 21, and 20 components, LC/ESI-MS shows 49, 50, and 46 components and LC/off-line MALDI shows 439, 286, and 157. Based on the different ionization methods and sample introduction procedures, it appears very likely that suppression effects as well as variation in sample concentration have lead to the observed disparity in the number of detected species. As mentioned previously, each fraction used for the offline MALDI

analysis was concentrated by a factor of 50 before mixing with the second layer of matrix.

Although direct MALDI is the most straightforward method of the three studied, it clearly suffers from an inability to detect the large number of components that are present in the cell extract. This is not surprising given the fact that many components in different relative amounts are analyzed simultaneously. In order to reduce the burden on the mass analyzer, a separation step prior to analysis is beneficial. The chromatographic process in LC/ESI-MS attenuates the problem of simultaneous detection of large numbers of species; however, it remains evident that the ionization process is still dealing with a mixture. Once a separation step has been included, the advantage that MALDI possesses over ESI for analysis of complex mixtures becomes evident. The chromatographic step prior to off-line MALDI analysis remains slower and more labor intensive than the chromatography coupled to ESI; however, it produces many more masses for inclusion in the proposed database.

The MALDI process generates fewer charge states than ESI, and as a result, MALDI is generally more amenable to mixture analysis. A typical ESI spectrum is made up of a molecular envelope consisting of several charge states and the resulting spectrum is usually difficult to interpret. This multiple charging effect was exploited in this work in order to produce a more confident set of peaks for deconvolution.

Figure 3-5 (A) shows a comparison of the offline MALDI spectrum of *E. coli* fraction 39 and the ESI spectrum of the corresponding peak observed during online LC/ESI-MS. Here MALDI reveals 15 components, while ESI results in six. The number of proteins detected in fraction 39 by LC/ESI-MS was atypical. As previously

mentioned, most fractions exhibited peaks from at most two or three species. Without further information from MS^n studies of the individual proteins, it remains a possibility that the spectra are composed of fragments from larger proteins. A good example is seen in Figure 3-5 (B), which contains peaks from a species at 13094, in addition to smaller signals from 6255 and 4066.

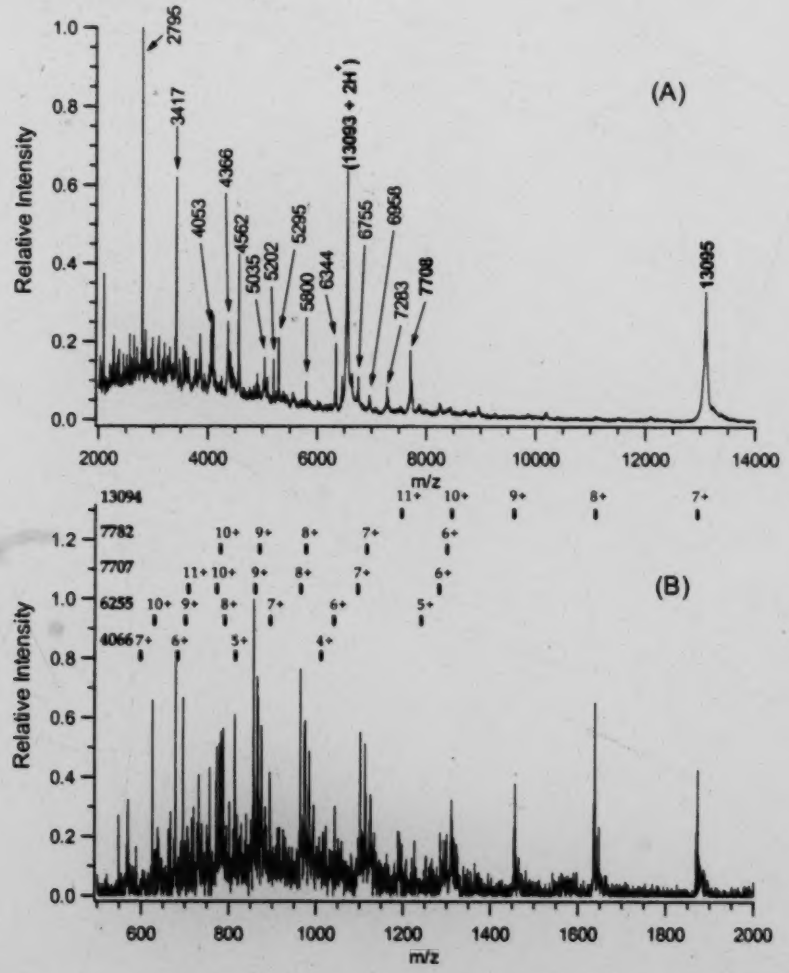

Figure 3-5 Comparison of species detected in fraction 39. (A) peaks observed by MALDI. (B) peaks seen by LC/ESI-MS. A total of six species were found; however, only five are shown for clarity.

Prior to off-line MALDI analysis, the collected HPLC fractions were pre-concentrated to ensure detection of the less abundant components. This pre-concentration step is likely partially responsible for the observation of more peaks by off-line MALDI than in LC/ESI-MS.

Generally, peaks produced by online separation and detection of the bacterial proteins were found to contain no more than three detectable components. The addition of a large amount of weak acid (e.g. acetic) proved necessary in order to observe enough protonated species. In the absence of acid, only a few characteristic protonated peaks were observable and their intensities were only just above the background.

3.3.3.1 Comparison of three data sets

Masses detected from direct MALDI are not all present in the mass tables produced by LC/off-line MALDI. These results suggest that LC/off-line MALDI is not ideal, and should not be regarded as the only method used for producing tables of biomarkers. Ion suppression remains an obstacle albeit not to the same extent as direct MALDI. For these complex bacterial extracts, it is quite common to find numerous components co-eluting during a one dimensional LC separation. The potential for protein loss and modification during sample work-up also remains an issue. To mitigate these effects, running samples prepared under different experimental conditions can be helpful. Variations in extraction solvent and contact time will undoubtedly extract a different set of proteins that may result in changes in the observed suppression effects.

Although in general, LC/off-line MALDI detects more peaks than LC/ESI-MS, the masses detected by LC/ESI-MS are not necessarily a subset of those produced by LC/off-line MALDI [15].

For *E. coli*, LC/ESI-MS produced 36 matches with the LC/off-line MALDI table. *B. megaterium* resulted in 27 of the 50 components matching the results found with LC/offline MALDI. *C. freundii* had 33 out 46 match the MALDI data. The results seem to indicate that sensitivity and suppression effects are different for the two techniques.

In the case of *B. megaterium*, many masses observed in direct MALDI are also detected in LC/ESI-MS. Alternatively, a few masses observed in direct MALDI are also detected in LC/ESI-MS for *E. coli* and *C. freundii*. Again this is a result of the different suppression processes in MALDI and ESI. The characteristics of the specific proteins existing in the three different bacteria account for the detection difference.

3.3.3.2 Database Creation

One of the objectives of this study was to evaluate whether a mass table created by LC/off-line MALDI is sufficient for bacteria identification. Although the off-line study produces more masses than the direct and online approaches, a comprehensive biomarker database is best composed of data from as many methods as possible. Ideally, such a database will contain only those biomarkers consistently observed with the various mass spectrometric methods. This approach eliminates the need for reproducibility, and only focuses on the masses of the chemotaxonomic markers. LC/ESI-MS is ideal for rapid analysis and allows for MS^n studies that LC/off-line

MALDI simply does not offer. In addition to this current work, a previously published report using the same *E. coli* strain but from a different batch [16], reveals a limitation of the current state of database formation. The proteins detected from the same strain, but different cultures, are not identical. Comparison of the off-line MALDI data shown in Table 3-8 to that produced in [16] reveals that 169 out of 307 peaks match. Environmental variations during cell growth in addition to sample preparation are likely causes for the observed differences. This leads to the question of the kinds of masses that should be included in the completed database. Combining the results from as many batches as possible would produce the most inclusive database; however, many masses are common to different strains and their presence cannot categorically be attributed to a single strain.

Once the tentative identity of the organism is established from the first round comparison of the initial biomarkers, tryptic and mass spectral fragmentation would provide the subsequent information to confidently establish a positive identification. Since the LC/ESI-MS approach does not lend itself well to performing tryptic digests, the MS^n capability of the ion trap would have to be utilized.

A comparison of the data produced during this experiment show one protein from *B. megaterium* (6345.5) and one from *C. freundii* (2431.5) in common with *E. coli*. There are no proteins in common between *B. megaterium* and *C. freundii*.

3.3.3.3 Comparison with Proteome Database

The current *E. coli* proteome database is quite extensive; however, Table 3-8 shows that a number of masses observed by MALDI are not listed in the proteome database (underlined masses are included in database). At the present time, however, the *E. coli* proteome database contains the most entries (2624 proteins in the range between 2 and 20 kDa), and as a result, any attempt at identification is inevitably biased towards *E. coli*. In this work, *B. megaterium* is an excellent example of the current limitation of the proteome database approach. Direct MALDI analysis showed 21 peaks, of which, only 4 matched with the *B. megaterium* proteome database (53 proteins in the range between 2 and 20 kDa). A search of another *Bacillus* strain, for example *subtilis* (1496 proteins in the range between 2 and 20 kDa), reveals 9 masses that match. In addition, it is not surprising to find that 12 of the 21 masses match with the *E. coli* proteome database.

The LC/ESI-MS data also produced masses matching those from other genus and species. For example for *B. megaterium*, 19 of 47 also match the *E. coli* database, 12 of 47 matched *B. subtilis*, however only 4 matched the *B. megaterium* database (with 0.02% accuracy).

It seems very probable that many of the observed proteins are from the fragmentation of larger proteins. In addition, post translational modifications as well as modifications during sample preparation can also account for the difference between the mass table and the database. Another possibility is a change in secondary structure resulting from the production of disulfide bonds, which would result in a mass 4 Da lower than that contained in the public mass database.

3.4 Conclusions

This chapter focused on the detection of low-mass peptides and proteins from bacterial cell extracts through the use of three techniques, namely direct MALDI, HPLC fractionation/off-line MALDI and LC/ESI-MS.

The methodology described in this chapter for development of a proteomic database for bacterial identification has shown that fraction collection followed by offline MALDI produces the most masses, followed by LC/ESI-MS and direct MALDI on the cell lysates. As a result, the most comprehensive database would be composed of data from all three of the methods presented.

Direct MALDI analysis reveals 20 to 29 protein components in the mass range from 2-20 kDa. HPLC fractionation and off-line MALDI-TOF analysis of individual fractions display 157 to 449 components. With an optimized HPLC condition for bacterial protein separation, LC/ESI-MS analysis of the extracts shows the number of detectable proteins in the range of 44 to 47. There are a number of commonly observed masses in the mass tables produced by the three methods for each bacterium; however, the differences observed are also quite substantial. In addition, for *E. coli*, where a large proteome database exists, a number of masses detected by the mass spectrometric methods are not present in the proteome database.

The proposed mass matching database approach to bacterial identification is limited by the size of the current protein mass database. Until more protein masses from organisms are catalogued, the use of mass spectrometry for identifying bacteria, will be biased towards bacteria with the greatest number of mass entries in the public databases.

3.5 Literature Cited

[1] Baar, B. L. M., *FEMS Microbiol. Rev.* **2000**, 24, 193-219.

[2] Evason, D. J.; Claydon, M. A.; Gordon, D. B. *Rapid Commun. Mass Spectrom.* **2000**, 669-672.

[3] Holland, R. D.; Duffy, C. R.; Rafii, F.; Sutherland, J. B.; Heinze, T. M.; Holder, C. L.; Voorhees, K. J.; Lay Jr, J. O. *Anal. Chem.* **1999**, 71, 3226-3230.

[4] Arnold, R. J.; Karty, J. A.; Ellington, A. D.; Reilly, J. P. *Anal. Chem.* **1999**, 71, 1990-1996.

[5] Krishnamurthy, T.; Ross, P. L. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1992-1996.

[6] Lynn, E. C.; Chung, M. C.; Tsai, W. C.; Han, C. C. *Rapid Commun. Mass Spectrom.* **1999**, 13, 2022-2027.

[7] Leenders, F.; Stein, T. H.; Kablitz, B.; Franke, P.; Vater, J. *Rapid Commun. Mass Spectrom.* **1999**, 13, 943-949.

[8] Claydon, M. A.; Davey, S. N.; Edwards-Jones, V.; Gordon, D. B. *Nature Biotech.* **1996**, 14, 1584-1586.

[9] Liang, X.; Zheng, K.; Qian, M. G.; Lubman, D. M. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1219-1226.

[10] Krishnamurthy, T.; Ross, P. L.; Rajamani, U. *Rapid Commun. Mass Spectrom.* **1996**, 10, 883-888.

[11] Wall, D. B.; Lubman, D. M.; Flynn, S. J. *Anal. Chem.* **1999**, 71, 3894-3900.

[12] Cain, T. C.; Lubman, D. M.; Weber Jr., W. J. *Rapid Commun. Mass Spectrom.* **1994**, 8, 1026-1030

[13] Chong, B. E.; Wall, D. B.; Lubman, D. M.; Flynn, S. J. *Rapid Commun. Mass Spectrom.* **1997**, 11, 1900-1908.

[14] van Adrichem J. H. M.; Bornsen, K. O.; Conzelmann, H.; Gass, M. A. S.; Eppenberger, H.; Kresbach, G. M.; Ehrat, M.; Leist, C. H. *Anal. Chem.* **1998**, 70, 923-930.

[15] Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R., *Rapid Commun. Mass Spectrom.* **1998**, 12, 456-464.

[16] Dai, Y.; Li, L.; Roser, D.C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1999**, 13, 73-78.

[17] Krishnamurthy, T.; Davis, M. T.; Stahl, D. C.; Lee, T. D. *Rapid Commun. Mass Spectrom.* **1999**, 13, 39-49.

[18] Dalluge, J. L.; Prasad, R. *BioTechniques* **2000**, 28,156-160.

[19] MacNair, J. E.; Opiteck, G. J.; Jorgenson, J. W.; Moseley, M. A. *Rapid Commun. Mass Spectrom.* **1997**, 11, 1279-1285.

[20] Liang, X., Zheng, K., Qian, M. G., Lubman, D. M. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1219-1226.

[21] Mauri, P. L.; Pietta, P. G.; Maggioni, A.; Cerquetti, M.; Sebastianelli, A.; Mastrantonio, P. *Rapid Commun. Mass Spectrom.* **1999**, 13, 695-703.

[22] http://expasy.cbr.nrc.ca/srs5/

[23] Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, 67, 1950-

[24] Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, 71, 1087-1091.

[25] Kuhlmann, F. E.; Apffel, A.; Fischer, S.; Goldberg, G.; Goodley, P. C. *J. Am. Chem. Soc.* **1995**, 6, 1221-1225.

CHAPTER 4

MALDI-TOF MS and LC/ESI-MS Analysis of *E. coli* 9637 Harvested at Different Growth Times

4.1 Introduction

Chapter 3 examined how m/z ratios obtained from bacterial extracts, when analyzed by three different mass spectral methods, can be used for producing a database consisting of m/z ratios. This chapter will focus on assessing the extent of variability in mass spectral data when one variable in the culturing process is varied. While there are a number of growth conditions such as media and nutrient concentration that can vary during cell growth, this chapter will focus on the effect of growth time.

A number of factors concerning the use of mass spectrometry as a tool for bacterial identification have been identified. Mass spectral reproducibility [1,2] as well as spectral differences within species and strains [3-6] have been reported in the literature.

Previous work by Reilly and co-workers examined the effect of growth time on MALDI spectra of *E. coli* 9637 whole cells [7]. Their work showed that the bacterial growth period has an affect on the MALDI spectra. Large variations in MALDI spectra were observed over certain windows during the growth period (e.g. 6-10 h and 22-30 h). Although the data proves that MALDI spectra are affected by variations in growth time, no data was presented comparing the MALDI results to LC/ESI-MS.

Work presented in this chapter will compare results from both MALDI and LC/ESI-MS analysis of solutions containing cellular extracts from *E. coli* ATCC 9637

to examine if the detection method used has any effect on the variation of mass spectra from samples prepared at different growth times.

4.2 Experimental

4.2.1 Materials

E. coli American Type Culture Collection (ATCC) 9637 samples cultured at the Edgewood RDE Center at Aberdeen Proving Ground, MD, were harvested at 6, 8, 10, and 12 hours after inoculation. The MALDI matrix, α-cyano-4-hydroxycinnamic acid (HCCA) and spectrophotometric grade trifluoroacetic acid (TFA) were purchased from Sigma Aldrich Canada (Oakville, ON). HPLC grade methanol was from Fisher Scientific Canada (Edmonton, AB). Water was obtained from a Milli-Q Plus purification system (Millipore Corporation, Bedford, MA, USA).

4.2.2 Sample Preparation

Crude water-soluble protein extracts were obtained using the same method used in Chapter 3. Approximately 5 –7 mg of lyophilized bacteria was weighed into a 1.5 mL siliconized vial. A 1 mL solution of 0.1% (v/v) aqueous TFA was then added to the solid bacteria and the mixture was vortexed and centrifuged. A small portion of the supernatant was removed and saved for MALDI analysis. A total of 5 extractions were performed, producing a total volume of 5 to 6 ml. The supernatants from each of the 4 bacterial growth times were combined and filtered with Microcon 3000 Da molecular weight cut-off filters (Amicon, Oakville, Ontario). The final filtered volume of 500 μL was then refrigerated for later analysis.

4.3.3 Instrumentation

LC/ESI-MS

Solvent delivery and separations were performed on an Agilent (Palo Alto, CA, USA) 1100 series HPLC equipped with an autosampler. All connections were made with 0.127 mm i.d. PEEK tubing and finger tight fittings (Upchurch Scientific). Chromatographic analysis was performed with a reversed phase Vydac C_8 column (2.1 mm i.d. x 150 mm; particle size, 5 μm; pore size, 300Å). Gradient elution was performed with solvent A (0.05% v/v aqueous TFA) and B (0.05% v/v TFA in acetonitrile). The following gradient profile was used (min:%B), 0:2, 10:20, 40:40, 45:55, 60:75, 75:2 at 200 μl/min. 30 μL of filtered, concentrated extract was injected for each growth period.

The HPLC effluent was analyzed with an Agilent 1100 series MSD quadrupole mass spectrometer. Agilent ChemStation software provided control of both the HPLC and mass spectrometer. The Igor Pro software package (WaveMetrics, Inc., Lake Oswego, OR) was used to reprocess the acquired data.

Post-column addition of glacial acetic acid was delivered at 100 μl/min in order to overwhelm the signal suppressing properties of the trifluoroacetic acid (TFA) in the mobile phase. The acid was added to the column effluent using a PEEK "Y" (Upchurch) connector and a syringe pump (Cole Parmer). The "Y" was connected to the electrospray interface with 30 cm of PEEK tubing. The length of the tubing exiting the "Y" and entering the ESI source was crucial for adequate mixing of the acid with the column effluent.

A variable fragmentation voltage was used during the ESI detection. The optimized voltage ramp was found to be: m/z 500 = 60V, 1000 = 120V, 3000 = 220V. In addition, the following parameters were used for MS detection of the chromatographic effluent: gain, 5; step size, 0.15; drying gas, 10 L/min; drying gas temperature, 350°C; nebulizer pressure, 25 psi; capillary voltage, 4500 V.

MALDI-TOF MS

MALDI experiments were performed on an Applied Biosystems Voyager Elite laser desorption/ionization time-of-flight mass spectrometer (Framingham, MA). Desorbed proteins ions were produced with a 337 nm pulsed nitrogen laser and detected in the linear mode of operation.

A two-layer sample preparation method with HCCA as the matrix was used for MALDI analysis. In the two-layer method [8], the first layer is formed by applying 1 μL of 100 mM HCCA in methanol/acetone (3/7 v/v) to the MALDI probe tip and allowing it to dry very quickly in air. For the second layer, sample solution was mixed 1:2 with saturated HCCA in formic acid/isopropanol/water/ (1/2/3 v/v/v); about 0.6-1 μL of the second layer solution is then applied onto the first layer and allowed to dry. The sample is then washed with 1 μL of distilled water for 10 seconds and then dried with a stream of nitrogen.

4.3 Results and Discussion

4.3.1 MALDI

Figure 4-1 shows the MALDI spectra for each of 5 successive extractions from a single portion of the lyophilized sample. The number of visible protein peaks

diminishes as the number of extractions increases. This suggests that extractions beyond 4 replicates are of little benefit in terms of increasing the amount of proteins in solution.

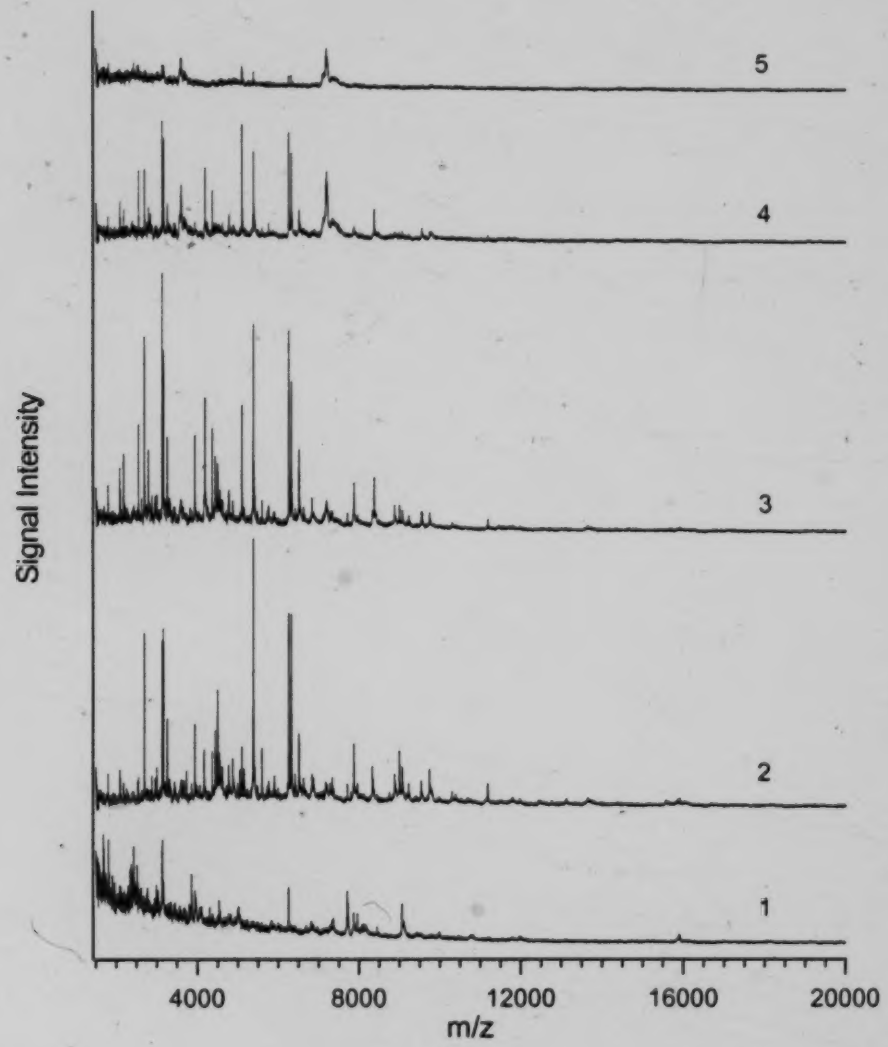

Figure 4-1 Extraction efficiency shown here with MALDI mass spectra of *E. coli* 9637 (10 hr. incubation) extracts. The first 1 mL extract is shown on the bottom, followed by each successive 1 mL extraction.

MALDI mass spectra of the combined and filtered cell extracts at different incubation times are shown in Figure 4-2. Even though identical strains of *E. coli* were analyzed, there is a substantial variation in the number of peaks observed.

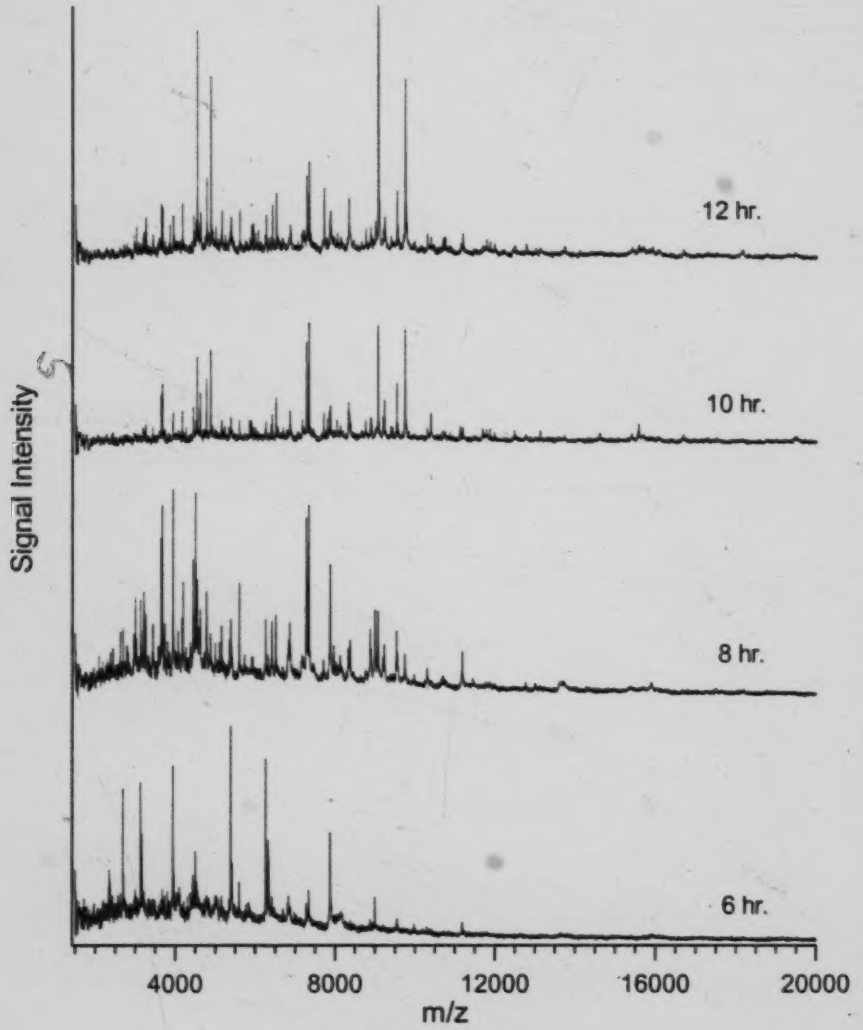

Figure 4-2 MALDI mass spectra of *E. coli* 9637 bacterial extracts at different incubation times. The concentrated raw extracts were mixed 1:2 with saturated HCCA in formic acid / methanol / water (1/2/3 v/v/v).

In particular, there is a marked difference between the 6-hour growth period, and the 8-, 10-, or 12-hour samples. However, the spectra of the 10-hour and 12-hour samples exhibit similar characteristics. The protein masses obtained from MALDI as shown in Tables 4-1 to 4-4 suggest that there are extensive variations among the samples, but the extent of variation is greater for the 6-hour sample compared to the others. This finding is consistent with that observed by Reilly and coworkers [9], who showed MALDI spectral variations over growth times ranging from 6 to 84 hours. The results presented in this chapter do not cover such a large growth period; however, they do show corroborate the expected variation in spectra according to growth time.

Table 4-1 MALDI $(M+H)^+$ 6 h incubation

2353.5	6255.6	7273.3	9549.9
2373.3	6316.6	7333.1	11183
5381.9	6331.8	7870.5	
5594.1	6837.8	8993.7	

Table 4-2 MALDI $(M+H)^+$ 8 h incubation

2623.9	6505.2	8870.7	11179
2959.1	6855.6	8988.9	12760
2998.8	7268.9	9059.3	13685
3205.9	7329.3	9221.5	15885
3729.0	7866.9	9530.5	
5379.8	7954.9	9735.4	
6252.7	8366.2	10294	

Table 4-3 MALDI $(M+H)^+$ 10 h incubation

3935.3	7329.6	9220.3	12479
4447.8	7701.2	9530.5	131120
4595.3	7843.0	9734.1	14597
5379.2	7865.9	10295	15571
5888.8	8032.3	10382	16678
6253.8	8321.3	11122	19499
6409.3	8750.3	11178	
6856.4	8871.6	11680	
7268.7	9058.6	11855	

Table 4-4 MALDI $(M+H)^+$ 12 h incubation

3854.0	6506.3	9059.9	11776
4162.6	6409.2	9531.9	11860
4988.9	7330.2	9735.1	11972
5592.3	7865.0	10294	16680
7270.4	8322.7	10688	18155
7704.0	8750.6	10734	
6254.5	8871.6	11180	

4.3.2 LC/ESI-MS

There is significantly less variation in LC/ESI-MS data than in the results obtained with MALDI. Figure 4-3 compares the total ion current (TIC) traces of the four different *E. coli* growth times. Aside from the 6-hour sample, the remaining three chromatograms show very similar patterns. What is more remarkable is that the protein masses obtained by LC/ESI-MS for the 8-, 10- and 12-hour samples are almost identical. In each of these samples, 42 or 43 protein masses are obtained, compared to 25, 33 and 26 masses observed in MALDI. The increased number of masses observed with LC/ESI-MS can be attributed to a reduction in ion suppression, which was also observed in Chapter 3. In addition, the chromatograms in Figure 4-4 produced by UV detection, exhibit similar characteristics as their TIC counterparts.

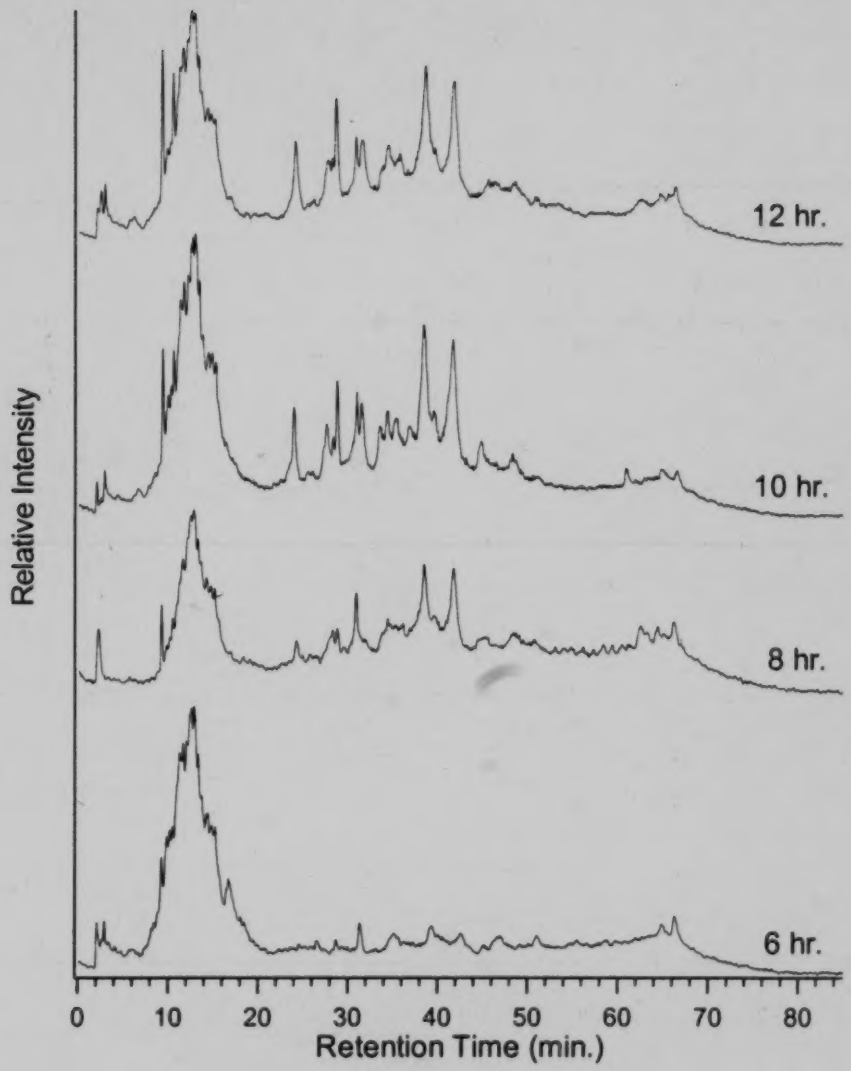

Figure 4-3 Total ion chromatograms of *E. coli* ATCC 9637 bacterial extracts at different incubation times. Glacial acetic acid was added post-column at 100 μL/min.

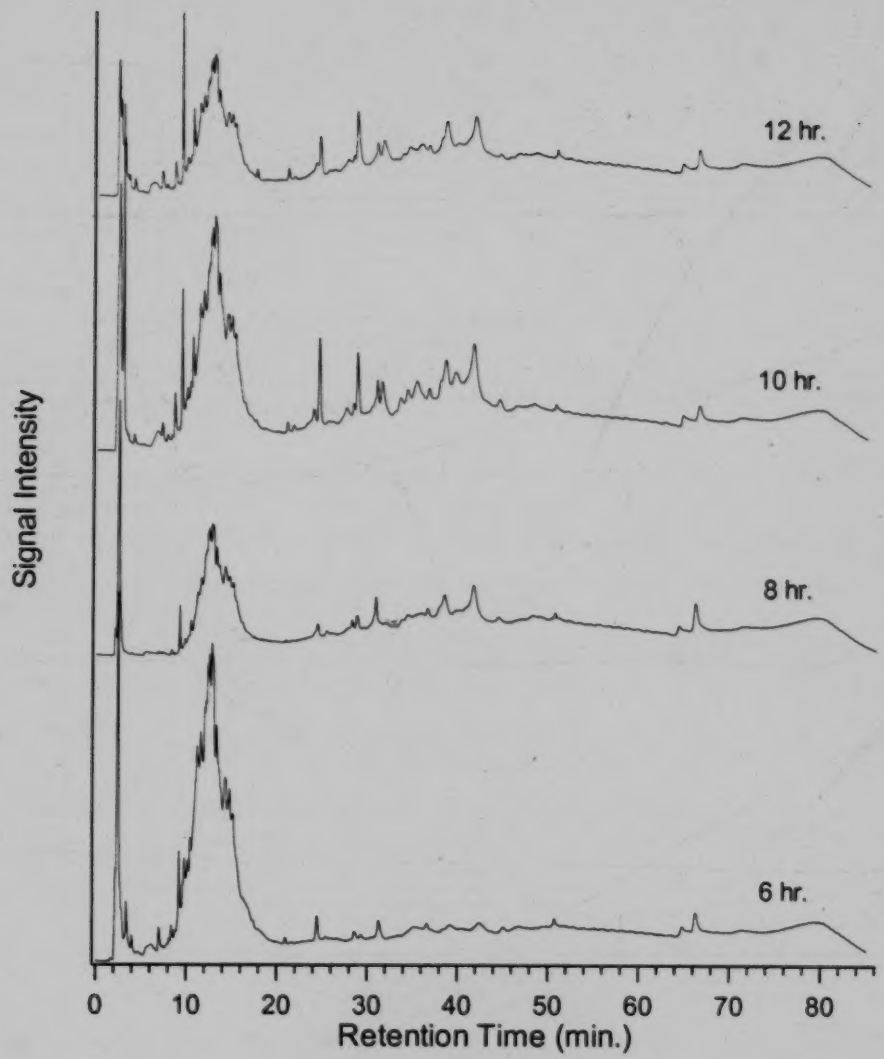

Figure 4-4 UV chromatograms at 214 nm of *E. coli* 9637 extracts at different incubation times. Both the UV and MS detectors were connected in series with the effluent passing first through the UV and then the MS detector.

It appears that as long as the cells are grown past the initial lag or adjustment phase (> 8 hours for *E. coli* judging from the data shown), reproducible LC/ESI-MS mass data for samples with different growth times can be achieved. However, more comparative studies employing different bacteria samples and longer growth times are warranted in order to validate the findings shown in this chapter.

In Chapter 3, LC/ESI-MS was used to analyze a different batch of the same strain of *E. coli* studied in this chapter. A total of 44 masses were found and are listed in Table 3-2. A comparison of the LC/ESI-MS masses determined in Chapter 3 with those presented in this chapter (see Tables 4-6, 7 and 8, representing incubation times of 8, 10 and 12 hours) shows 14 matching protein masses. However, only 1 mass matches *B. megaterium* (Table 3-3) and 4 masses match *C. freundii* (Table 3-4). This seems to indicate that while there is a variation in LC/ESI-MS data from sample to sample, a portion of the protein masses match with each other from the same bacteria and the number of masses matched is much greater than with those from different bacteria. In order to confirm if this is generally true, a larger database is needed. If so, further study into the subject could develop protein mass detection by LC/ESI-MS alone into an extremely powerful tool for bacteria identification.

Table 4-5 LC/ESI-MS $(M+H)^+$ 6 h incubation

2331.5	5020.5	7333.5	11781
2372.5	5381.2	7870.7	15693
2982.1	6316.2	9226.6	17516
3793.2	6411.2	9536.0	
4364.5	7272.2	11186	

Table 4-6 LC/ESI-MS $(M+H)^+$ 8 h incubation

2431.1	6411.3	8877.2	9740.1	14102
2472.4	6414.7	8993.2	10300	15410
2982.1	6857.4	9063.8	11124	15694
4364.1	7271.6	9191.4	11187	16685
5020.1	7274.9	9226.7	11685	17517
5096.6	7333.5	9421.6	11781	18163
5381.7	7707.6	9536.2	11978	18774
6255.4	7869.0	9541.4	12770	
6316.3	8326.2	9553.5	13131	

Table 4-7 LC/ESI-MS $(M+H)^+$ 10 h incubation

2431.1	6411.3	8878.2	9739.7	14102
2473.1	6414.8	8993.8	10300	15409
2982.0	6856.4	9064.2	11124	15693
4364.6	7271.7	9191.3	11186	16687
5020.0	7274.5	9226.6	11688	17515
5096.8	7333.2	9423.1	11781	18163
5381.1	7707.8	9536.3	11977	18773
6255.2	7869.3	9540.8	12771	
6316.1	8326.2	9553.9	13128	

Table 4-8 LC/ESI-MS $(M+H)^+$ 12 h incubation

2431.1	6255.0	7333.4	9191.6	11125	14102
2473.1	6316.3	7707.8	9226.8	11186	15409
2982.1	6411.4	7868.1	9422.6	11689	15693
4363.9	6414.5	8326.1	9536.4	11782	16687
5019.9	6856.4	8876.4	9541.4	11978	17515
5096.6	7271.8	8993.9	9740.1	12771	18162
5381.2	7274.7	9064.3	10300	13129	18775

In order to observe the effect of the acetic acid concentration in the post-column addition, various concentrations were chosen and the resulting ion chromatograms are shown in Figure 4-5. There is very little variation in terms of the maximum ion current observed when infusing concentrations up to 90% (v/v) acetic acid. However, infusing glacial acetic acid produces a significantly larger ion signal. The larger total ion current is also observed in Figure 4-6, where individual mass spectra at an elution time of 38 minutes are shown for each acid concentration.

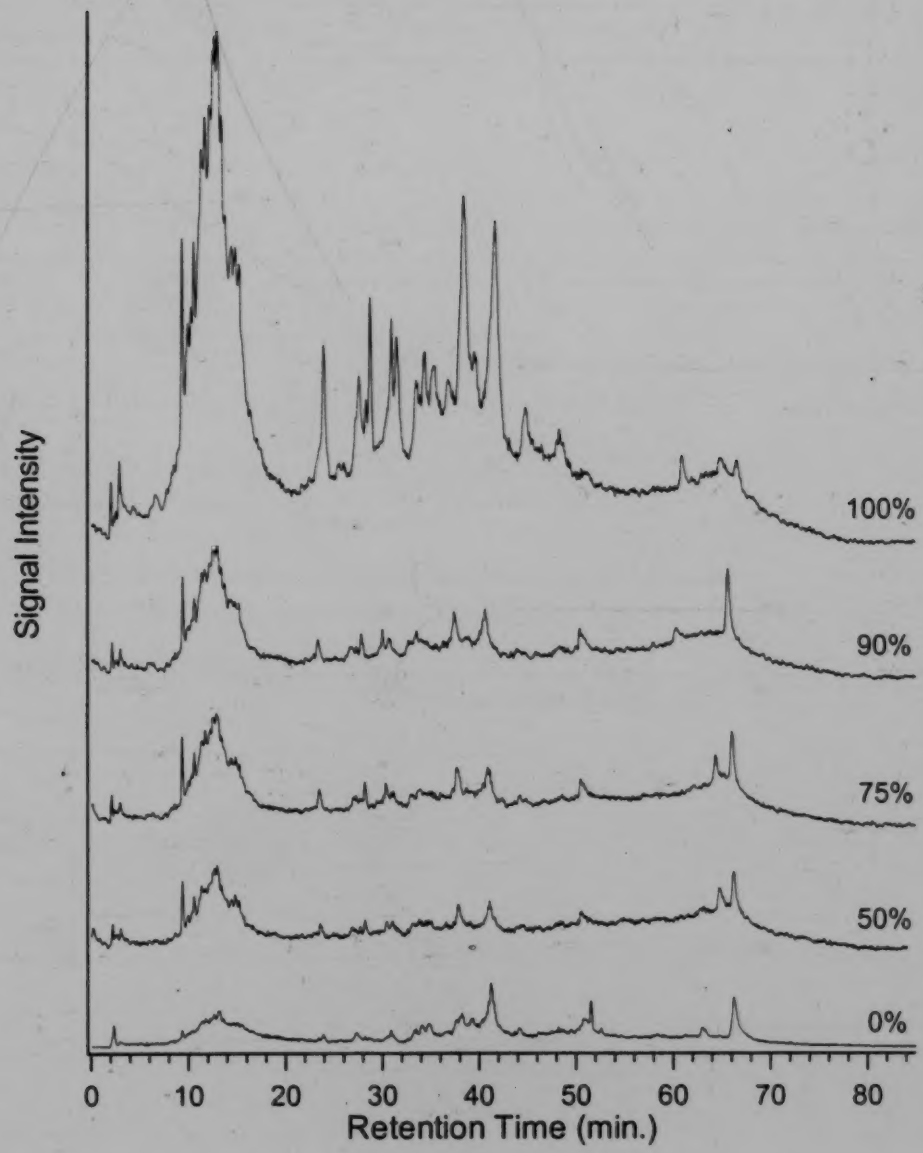

Figure 4-5 Total ion chromatograms of *E. coli* 9637 (10 hr. incubation) at varying concentrations of post-column acetic acid.

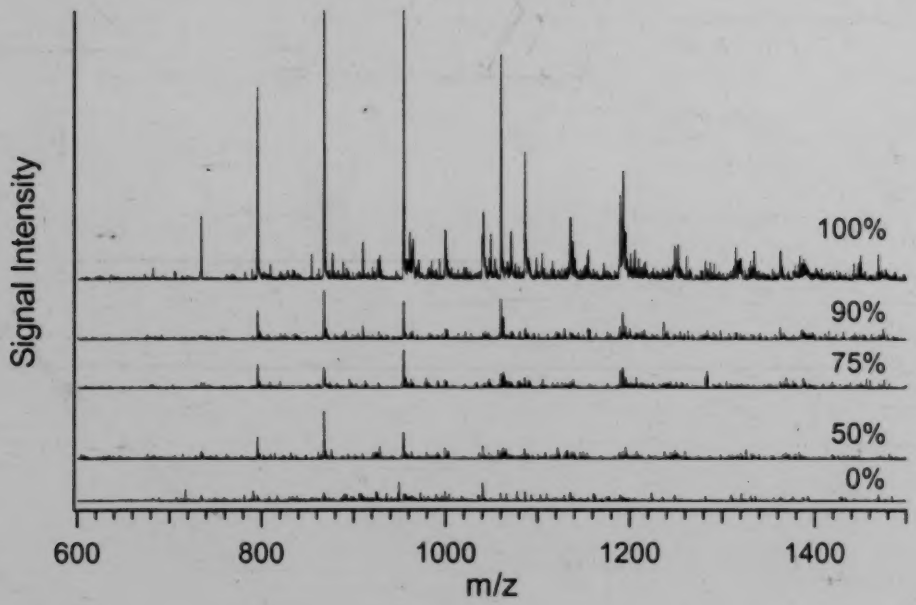

Figure 4-6 Mass spectral signal enhancement observed upon addition of post column acetic acid at the listed concentrations. Acid was added at 100 μL/min, or one half the mobile phase flow rate.

4.4 Conclusions

Bacteria cultured for different lengths of time exhibit variations in terms of the number of protein mass detected by MALDI. Some peaks show strong intensities for one culture time, while at other times they are greatly diminished or completely absent. In LC/ESI-MS, the 6-hour sample produces 18 proteins (compared to 14 by MALDI). For the 8, 10 and 12-hour samples, 42 or 43 proteins are detected (compared to 25, 33 and 26 in MALDI). The LC/ESI-MS data exhibits much less variability in terms of the masses detected from the 8, 10 and 12-hour cultures.

The masses detected by LC/ESI-MS are almost identical for the samples grown for 8, 10 and 12 hours. Based on these results, LC/ESI-MS exhibits less spectral variability than MALDI for analyzing samples taken from cultures grown at different times. This observation requires further study in order to provide more conclusive evidence; however, if shown to be true, it will have a significant implication on the overall approach to bacterial identification.

4.5 Literature Cited

[1] Liang, X.; Zheng, K.; Qian, M. G.; Lubman, D. M. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1219-1226.

[2] Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, 12, 456-464.

[3] Krishnamurthy, T.; Ross, P. L.; Rajamani, U. *Rapid Commun. Mass Spectrom.* **1996**, 10, 883-888.

[4] Holland, R. D.; Wilkes, J. G.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O., Jr. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1227-1232.

[5] Claydon, M. A.; Darey, S. N.; Edwards-Jones, V.; Gordon, D. B. *Nature Biotechnol.* **1996**, 14, 1584-1586.

[6] Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1998**, 12, 630-636.

[7] Arnold, R. J.; Karty, J. A.; Ellington, A. D.; Reilly, J. P.; *Anal. Chem.* **1999**, 71, 1990-1996.

[8] Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, 71, 1087-1091.

CHAPTER 5

Automated Analysis of LC/ESI-MS Data from Bacterial Extracts

5.1 Introduction

Although liquid chromatography/electrospray ionization mass spectrometery (LC/ESI-MS) is able to rapidly produce data in terms of a chromatogram, there is little information gained from the chromatogram alone. For the production of mass tables, it is necessary to average the mass spectra produced over a selected region of the chromatogram. It is this portion of the analysis that requires the greatest amount of user intervention, not only in time, but also in interpretation of the charge states making up the deconvoluted m/z. To date, there has been very little work done on exploring ways of completely automating the data analysis portion of an LC/ESI-MS experiment.

One study examined both manual and automated production of deconvoluted molecular masses produced from LC/ESI-MS analysis of human plasma filtrates [1]. The automated analysis was considered a helpful aspect of the overall study due to the complex nature of the analyte, and the resulting multifaceted nature of the chromatogram. The approach taken however utilizes a third party program to analyze and interpret the data. This program is not available for analyzing data produced with the ChemStation software bundled with Agilent chemical analysis products.

In order to facilitate the analysis of complex mixtures, chromatographic resolution may be enhanced through mathematical handling of the chromatographic data [2]. In addition, overlapping chromatographic peaks can be separated through

examination of peak centroids [3] as well as extraction of mass spectra from poorly separated chromatographic peaks [4,5].

A novel method for rapid production of protein mass tables from LC/ESI-MS of bacterial cell lysates has been developed. The approach relies on the production of an oscillating signal in the total ion chromatogram (TIC) and subsequent analysis by the auto-integration feature found in Agilent's ChemStation data analysis software. The artificial chromatographic peaks are produced when the static mixer on the Agilent 1100 series pump is removed. The reduced mixing is believed to produce sections richer in organic content, which subsequently produce an enhanced signal when proteins are sprayed. Using this method, the TIC is divided into a series of narrow peaks with a baseline width of approximately one minute. In this work, the technique is shown to work favorably for two bacterial extracts, namely *Escherichia coli* (*E. coli)* ATCC 9637 and *Bacillus globigii* (*B. globigii*). Results from manual and automated data analysis are compared and the method's application to bacterial identification is discussed. The automated approach detects a slightly smaller subset of peaks than if the data were completely examined manually; however, the amount of time saved through automation makes the approach appealing for rapid production of mass tables for comparison with protein mass databases.

5.2 Experimental

5.2.1 Materials

Escherichia coli (ATCC 9637) and *B. globigii* were a gift from the Edgewood RDE Center at Aberdeen Proving Ground, MD. Spectrophotometric grade

trifluoroacetic acid (TFA) was purchased from Sigma Aldrich Canada (Oakville, ON). HPLC grade acetonitrile and glacial acetic acid were from Fisher Scientific Canada (Edmonton, AB). Water was obtained from a Milli-Q Plus purification system (Millipore Corporation, Bedford, MA, USA).

5.2.2 Sample Preparation

Bacterial extracts were prepared by a solvent suspension method. About 7 mg of lyophilized bacteria was suspended in 1 ml 0.1% (v/v) aqueous TFA, vortexed for 3-5 minutes and centrifuged. A total of 5 extractions were performed, producing a total volume of 5 ml. The supernatants from each of the bacterial samples were combined and filtered by Microcon-3 filters with a molecular weight cut-off of 3000 Da (Amicon, Oakville, Ontario).

5.2.3 Instrumentation

LC/ESI-MS

Solvent delivery and separations were performed on an Agilent (Palo Alto, CA, USA) 1100 series HPLC equipped with an autosampler. All connections were made with 0.127 mm i.d. PEEK tubing and fingertight fittings (Upchurch Scientific). Chromatographic analysis was performed with a reversed phase Vydac C_8 column (2.1 mm i.d. x 150 mm; particle size, 5 μm; pore size, 300Å). A flow rate of 200 μl/min was used for all samples in the LC/ESI-MS analysis. Gradient elution was performed with solvent A (0.05% v/v aqueous TFA) and B (0.05% v/v TFA in aceonitrile). The

following gradient profile was used (min:%B), 0:2, 10:20, 40:40, 45:55, 60:75, 75:2 at 200 µl/min. Each HPLC run consisted of a 30 µL injection of bacterial extract.

The HPLC effluent was analyzed with an Agilent 1100 series MSD quadrupole mass spectrometer. Agilent ChemStation software provided control of both the HPLC and mass spectrometer. The Igor Pro software package (WaveMetrics, Inc., Lake Oswego, OR) was used to reprocess the acquired data.

Post-column addition of glacial acetic acid was delivered at 100 µl/min in order to overwhelm the signal suppressing properties of the trifluoroacetic acid (TFA) in the mobile phase. The acid was added to the column effluent using a PEEK "Y" (Upchurch) connector and a syringe pump (Cole Parmer). The "Y" was connected to the electrospray interface with 30 cm of PEEK tubing. The length of the tubing exiting the "Y" and entering the ESI source was crucial for adequate mixing of the acid with the column effluent.

A variable fragmentation voltage was used during the ESI detection. The optimized voltage ramp was found to be: m/z 500 = 60V, 1000 = 120V, 3000 = 220V. In addition, the following parameters were used for MS detection of the chromatographic eluent: gain, 5; step size, 0.15; drying gas, 10 L/min; drying gas temperature, 350°C; nebulizer pressure, 25 psi; capillary voltage, 4500 V.

5.3 Results and Discussion

The goal of this work is to provide a simple and effective way to modify the chromatographic output of an LC/ESI-MS run in order to expedite the data analysis process. The current ChemStation analysis software provided by Agilent Technologies does not provide the user the opportunity to average the acquired mass spectra over a constant time frame. The software is programmed to identify actual peaks in a chromatogram. In cases where small perturbations in the signal exist, the program can interpret these as peaks eventhough the peak width may only be a few seconds. In such instances, the program identifies too many peaks and the integration is inefficient.

A thorough analysis of the chromatograms shown in this thesis reveals several proteins eluting in regions where no peaks are present (i.e. the baseline). In this case, the automatic peak finder would skip over what it would classify as background. However, if a peak were introduced in a region where it previously did not exist, the mass spectra could then be averaged and the resulting spectrum deconvoluted. In this manner, the chromatographic information can be used much more thoroughly, in effect; the entire chromatogram can now be searched and deconvoluted automatically. Figure 5-1 shows how the removal of the static mixer influences the pattern of the TIC.

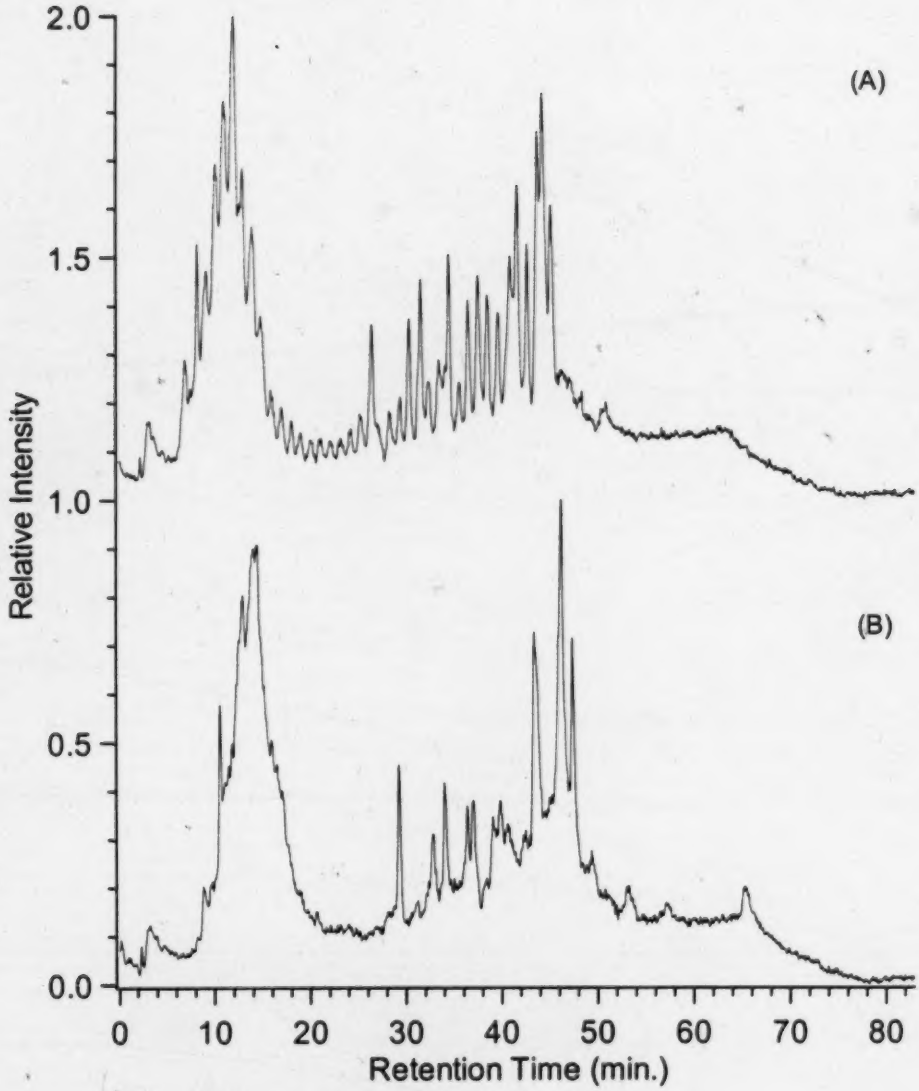

Figure 5-1 TIC of *E. coli* 9637 cell extract with post-column glacial acetic acid.
(A) Static mixer removed
(B) Static mixer installed

5.3.1 Automated peak identification.

Removal of the static mixer from the flow path of the HPLC pumps produced an oscillating TIC signal. The source of the relatively evenly spaced peaks is thought to arise from the production of segmented flow in which the ionization process is augmented and diminished as the segments are sprayed. The effect was only observed as proteins eluted from the column. The artificial peaks produced through this process,were used to help the ChemStation program identify peaks in the chromatogram. The ChemStation program does not possess the capability to average spectra over a preset time window for the entire chromatogram. Although this may be possible in future releases, this current deficiency makes it difficult to automatically find peaks, average the mass spectra and then produce a list of deconvoluted masses.

5.3.1.1 Manual data interpretation.

The peaks found by ChemStation through the automatic peak identification process had peak widths at their base of about 60 seconds. Each mass spectrum contained in the peak was averaged and background subtracted. The resulting mass spectra were then deconvoluted in order to determine the $(M+H)^+$ mass for inclusion in the mass tables. The shapes of the proposed charge profiles were then examined manually to see if they matched the expected Gaussian profile produced by the electrospray process.

5.3.1.2 Automated data interpretation.

Instead of manually interpreting the proposed deconvoluted masses, an attempt was made to completely automate the process. In this case, the first mass proposed by the program was taken. It was found that the second mass in the list was unlikely to be correct, and as a result, this mass was not used to compile the mass tables found in this paper. In order to compare the results with the manual interpretation, identical parameters were used in the deconvolution program. Table 5-1 summarizes the masses found for both the manual and automated interpretation of the deconvoluted mass spectra. The automated data interpretation was very close to the manual approach, finding 30 masses compared to 32 in the manual case. In addition, the effect of a larger noise cutoff was examined. If the noise cutoff is increased from 2000 to 5000, only 25 masses are found, whereas if the cutoff is increased further to 10000, only 15 masses are found. The results shown in Table 5-1 show that 70% of the masses detected with the automated interpretation of the data were confirmed by manually examining the same data.

Manual Interpretation	Automated Interpretation	Automated Interpretation	Automated Interpretation
2982.0	2982.0	2982.0	5096.4
4465.1	5096.6	5096.6	7707.5
5096.5	6052.8	6052.8	7904.7
6255.3	6255.3	6255.3	8510.2
6316.0	6315.5	7090.4	9064.1
6411.2	7090.4	7707.5	9226.6
6856.7	7707.5	7904.7	9536.1
7273.4	7867.5	9064.1	9739.8
7333.0	8325.8	9191.4	10591
7707.5	9064.1	9226.6	11044
7867.5	9191.4	9292.1	11126
8322.6	9226.6	9236.1	11782
8325.9	9427.0	9739.8	11977
8834.8	9536.1	10300	13128
8876.7	9739.8	10772	18162
9064.1	10300	11044	
9190.5	10772	11126	
9226.6	11125	11186	
9387.4	11186	11435	
9536.1	11689	11689	
9540.4	11782	11782	
9739.8	11977	11977	
10299	13128	13128	
11124	13421	15409	
11186	14867	18162	
11689	15409		
11782	16064		
11976	16334		
13128	18162		
15410	18773		
15692			
18162			
Noise = 2000 counts	Noise = 2000 counts	Noise = 5000 counts	Noise = 10000 counts

Table 5-1 Automated peak identification. Masses found through both manual and automated interpretation of the deconvoluted mass spectra by ChemStation. The masses listed were found by deconvoluting mass spectra from peaks found automatically through removal of the static mixer. At the bottom of the columns are listed the number of counts in each mass spectrum below which potential peaks are rejected.

5.3.2 Manual peak identification.

Incorporating the static mixer into the HPLC setup produced a TIC without the exaggerated peak profiles. In this case, the TIC was studied by manually averaging spectra over a selected time scale. The effect in this case tried to mimic artificial peaks created when the static mixer is removed. Mass spectra were averaged over one minute intervals starting from the onset of a signal, as well as offsetting the averaging by 30 seconds and starting at 30 seconds and averaging over 60 seconds to 1.5 minutes. This approach was chosen because ChemStation does not currently have the capability of integrating over regions where no obvious peaks are present.

5.3.2.1 Manual data interpretation

The deconvoluted mass spectra were analyzed in the same manner as those produced with the automated peak identification. In each case a mass spectral profile was deemed reasonable if it displayed a Gaussian distribution and if the majority of the peaks were greater than three times the background. In addition, the retention times of the proposed masses were considered, and masses were discarded if very large proteins were eluting early on during the run.

The highest number of masses occurred when starting at zero minutes and averaging over 60 seconds. A total of 44 masses were found, of which 39 match those found using the other two methods of spectral averaging. If the time period averaged was made too wide (e.g. 90 seconds instead of 60 seconds), slightly fewer masses were found. The majority of the masses found using the larger window were also found

when the time period was offset; however, 8 masses were not present, and none of them matched the 44 found without offsetting the start time.

5.3.2.2 Automated data interpretation

In this case, automated interpretation of the charge profiles was undertaken on spectra averaged for 60 seconds starting at 0 minutes. The automated approach found 72% of the masses found by manual interpretation of the same data. The remaining 28% are considered false positives as their deconvoluted profiles are easily dismissed upon closer examination.

Table 5-2 summarizes the masses found for both manual and automated interpretation of peaks found by manually examining the TIC over fixed time intervals.

Manual Interpretation (0-1 min)	Manual Interpretation (0-1.5 min)	Manual Interpretation (0.5-1.5 min)	Automated Interpretation (0-1 min)
2430.1	2431.1	2430.9	2032.4
2982.0	2982.0	2982.1	2430.2
3027.2	3027.3	4363.4	2981.1
4363.8	4363.4	4464.6	2984.1
4464.0	4465.0	5020.8	5095.9
5005.0	5020.7	5096.6	6254.4
5020.6	5096.8	5381.7	6315.3
5096.9	5379.9	6255.6	6855.8
5381.8	6255.5	6316.4	7089.4
6255.4	6316.4	6414.6	7332.1
6316.4	6411.3	6856.9	7707.0
6414.9	6414.4	7274.6	7866.6
6856.8	6856.8	7333.2	8325.1
7272.6	7274.6	7707.7	8992.0
7274.3	7333.2	7722.2	9063.2
7333.1	7707.7	7847.2	9190.5
7637.6	7867.7	7868.6	9225.7
7708.0	8326.0	8326.0	9273.4
7722.8	8834.6	8751.9	9535.2
7867.9	8876.8	8834.6	9539.5
7870.4	8992.8	8876.3	9738.7
8326.0	9064.2	9064.2	11044
8877.2	9191.4	9191.5	11185
9064.2	9226.7	9226.7	11687
9191.1	9536.2	9428.1	11780
9226.8	9739.7	9536.2	12770
9536.2	9755.6	9540.3	12932
9540.0	10087	9739.8	14808
9634.9	11132	12300	15408
9739.8	11186	11186	15692
9755.7	11688	11688	16464
11124	11781	11781	16685
11133	12771	11977	
11186	13127	12484	
11688	15409	12772	
11781	15692	13128	
11977	16686	15409	
12770	18162	18162	
13128			
15409			
15693			
16686			
18163			
18772			

Table 5-2 Manual peak identification. Masses found through both manual and automated interpretation of the deconvoluted mass spectra by ChemStation. The masses listed were found by deconvoluting mass spectra from peaks found by averaging over the listed time frame.

The completely automated approach to producing mass tables can be accomplished through removal of the static mixer and using the first mass chosen by the ChemStation software. The results should be interpreted with caution though due to the complex nature of the spectra involved. The example with *E. coli* demonstrated that 70% of the masses found with automated peak identification and automated interpretation match those found with manual peak identification and manual interpretation. The majority of the 30% that did not match were found during the first 15 minutes of the run, and most possessed molecular weights greater than 10 kDa. If the retention time is also taken into account, thereby removing these larger proteins, then 84% of the masses found with complete automation match those found manually.

The results show little difference between masses obtained with the static mixer removed and those found by manually averaging spectra over a fixed time period. As a result, the problem addressed here only arises from an inability of the data analysis program to perform the required signal averaging. Future versions may have this capability; however, data analysis with current software should benefit from the production of more exaggerated peaks in the TIC.

A gram-positive bacterium was also chosen to exhibit the potential of the proposed method. The solvent suspension method used to extract the cell contents was not as effective at rupturing the thicker cell walls of *B. globigii*. As a result, the chromatograms exhibit a much weaker signal compared to that of *E. coli*. The effect of removing the static mixer however is quite apparent and can be seen in Figure 5-2.

Even though the signal from the gram-positive *B. globigii* is much flatter and weaker than *E. coli*, manual examination of the chromatogram found 27 masses, and

automated discovered 24. The proposed method is able to transform the chromatograms from both gram-negative and gram-positive cell extracts. The effects of the oscillating signal are only present when components elute from the column. This was verified by conducting blank gradient runs with the static mixer removed and with post-column glacial acetic acid. The electrospray current did not display the distinctive oscillations.

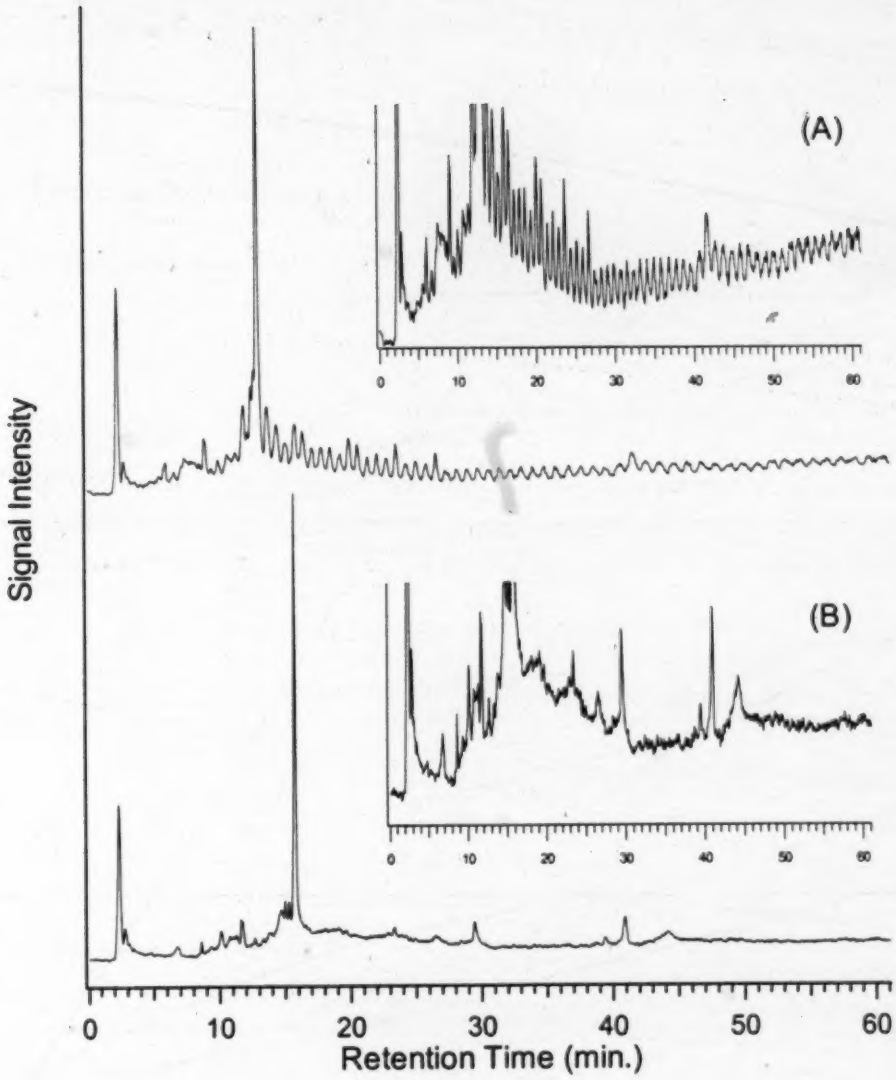

Figure 5-2 TIC of *B. globigii* cell extract with post-column glacial acetic acid.
(A) Static mixer removed
(B) Static mixer installed

Examples of some mass spectra involved in this work are shown in Figure 5-3. The upper trace (A) shows the output observed from a peak containing two components. It is not known if the second component is the result of fragmentation, or from a separate species altogether. The charge profiles correspond to protonated molecular weights of 11186 and 18162 Da. The second example displays the peaks associated with a protein with a protonated molecular weight of 7707.5 Da. In both cases, the Gaussian distribution of charge is clear and therefore likely to correspond to the masses listed.

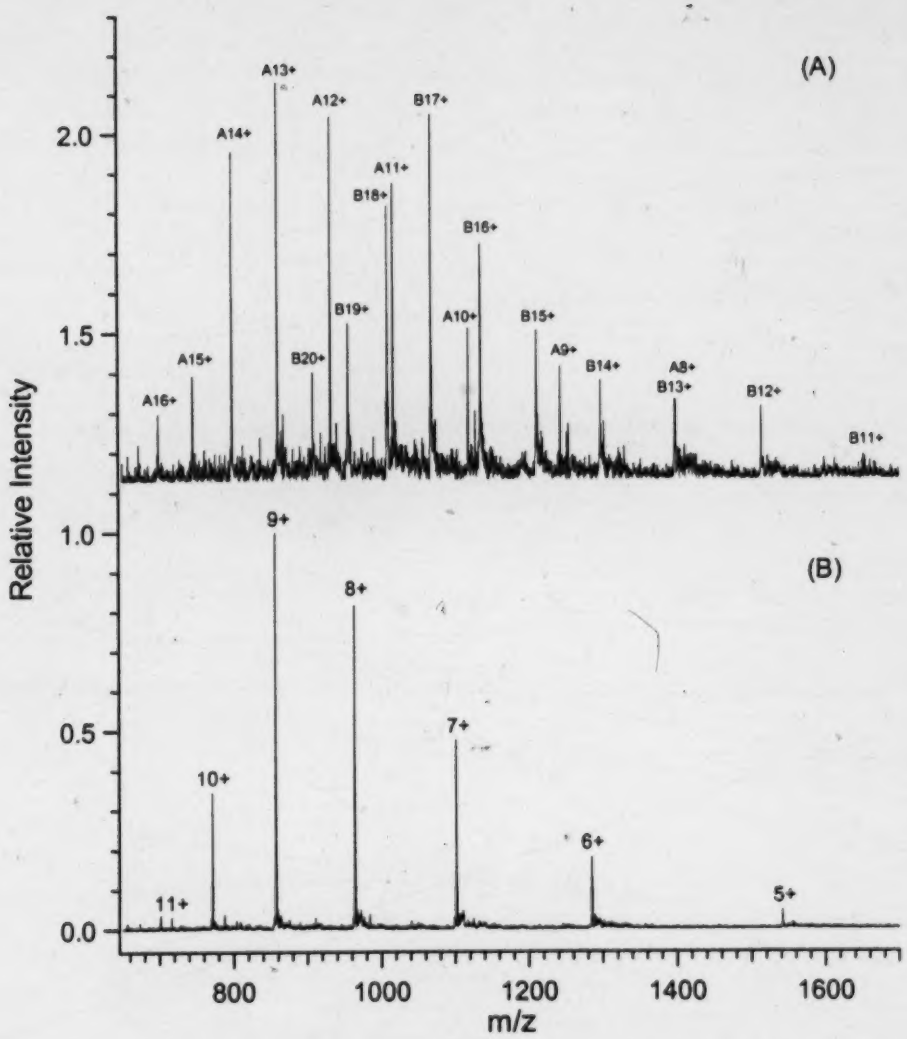

Figure 5-3 Example of mass spectra showing charge states observed from *E. coli* cell extracts. The majority of the peaks, both manually and automatically identified, contained more than one protein.

(A) Two components with overlapping charge envelopes

A. $(M+H)^+ = 11186$

B. $(M+H)^+ = 18162$

(B) One component with $(M+H)^+ = 7707.5$

5.4 Conclusions

Automated peak identification and data interpretation has been successfully applied to the analysis of total ion chromatograms of bacterial lysates. The proposed method layers an oscillating signal over the normal signal produced by the ionization process. The signal enhancement is believed to arise from the production of solvent plugs that augment the total ion current as proteins elute from the column. Automated peak identification and spectral interpretation found 30 masses, whereas manually interpreting the same data revealed 32 masses. Of the 30 masses obtained through completely automated means, 70% were confirmed through manual interpretation of the same spectra. The completely automated approach is rapid and capable of easily producing a set of masses in which the majority corresponds to those found by manual means. Other than the noise threshold, altering the search parameters had very little effect on which masses were detected. Choosing only the first mass of those suggested in the deconvolution table potentially underestimates the number that may be present; however, the next mass in the table rarely matched those found by manually analyzing the data.

Implementation of the proposed method allows for altering the mass spectral current without significantly varying the actual separation or the number of protein masses that are detected. Future updates to the software may include the capability to average spectra over a fixed time frame; however, until such versions are made available, the proposed method offers an alternative to the lengthy and operator dependent task of manually identifying peaks and interpreting the results.

5.5 Literature Cited

[1] Raida, M., Schultz-Knappe, P., Heine, G., Forssmann, W., *J. Am. Soc. Mass Spectrom.* **1999**, 20, 45-54.

[2] Pool, W. G.; de Leeuw, J. W.; van de Graaf, B. *J. Mass Spectrom.* **1997**, 32, 438-443.

[3] Colby, B. N. *J. Am. Soc. Mass Spectrom.* **1992**, 3 558-562.

[4] Ghosh, A.; Anderegg, R. J. *Anal. Chem.* **1989**, 61, 73-77.

[5] Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C.; Duffielk, A. M. *Anal. Chem* **1976**, 48, 1368-1375.

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions and Future Work

Several mass spectrometric methods for bacterial identification were explored. Chapter two examined some fundamental and technical issues regarding nanoelectrospray mass spectrometry. Nanoelectrospray emitters were fabricated and used for analysis of a tryptic digest of cytochrome c, in addition to an MS^2 study on fragmenting an unknown *E. coli* protein. A combination of three techniques, namely MALDI, HPLC fractionation/off-line MALDI and LC/ESI-MS were used in Chapter 3. The goal was to compare the three approaches and evaluate their potential for producing databases containing the molecular weights of bacterial proteins. The issue of bacterial growth time and its affect on MALDI and LC/ESI-MS spectra was explored in Chapter 4. The masses detected using LC/ESI-MS exhibited much less dependence on growth time than MALDI. A method that facilitates the analysis of complex LC/ESI-MS chromatograms was presented in Chapter 5. The proposed method layers an oscillating signal over the chromatogram. Implementation of the proposed method allows for altering the mass spectral current while facilitating the location and deconvolution of the mass spectral information that produced the peak. The proposed method offers an alternative to the lengthy and operator dependent task of manually identifying peaks and interpreting the results.

The characterization of bacteria with MALDI and LC/ESI-MS remains in the early stages of development. Further study in the LC/ESI-MS analysis is required using

different extraction conditions such as sonication, chemical and French press procedures. In addition, an expansion in the number of *E. coli* growth times to include data up to 48 or 72 hours is also warranted. A more complete examination of the fractionation and nanospray MS^2 using the quadrupole ion trap will also add to the basic mass data obtained from LC/ESI-MS and MALDI.